

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
6 October 2005 (06.10.2005)

PCT

(10) International Publication Number  
**WO 2005/091971 A2**

(51) International Patent Classification: Not classified

(21) International Application Number:  
PCT/US2005/007404

(22) International Filing Date: 3 March 2005 (03.03.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/550,023 3 March 2004 (03.03.2004) US

(71) Applicant (for all designated States except US): **ISIS PHARMACEUTICALS, INC.** [US/US]; 2292 Faraday Avenue, Carlsbad, CA 92008 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SAMPATH, Rangarajan** [US/US]; 12223 Mannix Road, San Diego, CA 92129 (US). **HALL, Thomas, A.** [US/US]; 5239 Wohlford Street, Oceanside, CA 92056 (US). **ESHOO, Mark, W.** [US/US]; 615 Glenmont Drive, Solana Beach, CA 92075 (US).

(74) Agent: **LEGAARD, Paul, K.**; Cozen O'Connor, 1900 Market Street, Philadelphia, PA 19103 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PI, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS FOR USE IN IDENTIFICATION OF ALPHAVIRUSES

(57) Abstract: The present invention provides oligonucleotide primers and compositions and kits containing the same for rapid identification of alphaviruses by amplification of a segment of viral nucleic acid followed by molecular mass analysis.



WO 2005/091971 A2

## COMPOSITIONS FOR USE IN IDENTIFICATION OF ALPHAVIRUSES

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Application Serial No. 60/550,023, filed March 3, 2004, which is incorporated herein by reference in its entirety.

### STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with United States Government support under DARPA/SPO contract BAA00-09. The United States Government may have certain rights in the invention.

### FIELD OF THE INVENTION

[0003] The present invention relates generally to the field of genetic identification and quantification of alphaviruses and provides methods, compositions and kits useful for this purpose, as well as others, when combined with molecular mass analysis.

### BACKGROUND OF THE INVENTION

#### A. Alphaviruses

[0004] Togaviridae is a family of viruses that includes the genus alphavirus. Alphaviruses are enveloped viruses with a linear, positive-sense single-stranded RNA genome. Members of the alphavirus genus include at least 30 species of arthropod-borne viruses, including Aura (AURA), Babanki (BAB), Barmah Forest (BF), Bebaru (BEB), Buggy Creek, Cabassou (CAB), Chikungunya (CHIK), Eastern equine encephalitis (EEE), Everglades (EVE), Fort Morgan (FM), Getah (GET), Highlands J (HJ), Kyzylagach (KYZ), Mayaro (MAY), Middelburg (MID), Mucambo (MUC), Ndumu (NDU), O'nyong-nyong (ONN), Pixuna (PIX), Ross River (RR), Sagiya (SAG), Salmon pancreas disease (SPDV), Semliki Forest (SF), Una (UNA), Venezuelan equine encephalitis (VEE), Western equine encephalitis (WEE) and Whataroa (WHA) virus ("The Springer Index of Viruses," pgs. 1148-1155, Tidona and Darai eds., 2001, Springer, New York; Strauss and Strauss, Microbiol. Rev., 1994, 58, 491-562). Alphaviruses are evolutionarily differentiated based on nucleotide sequence of the nonstructural proteins, of which there are four (nsP1, nsP2, nsP3 and nsP4). The genus segregates into New World (American) and Old World (Eurasian/African/Australasian) alphaviruses based on geographic distribution. It is estimated that New World and Old World viruses diverged between 2,000 and 3,000 years ago (Harley et al., Clin. Microbiol. Rev., 2001, 14, 909-932).

**[0005]** Among the alphavirus species, there are seven distinct serocomplexes (SF, EEE, MID, NDU, VEE, WEE and BFV) into which members of the genus are sub-divided (Khan et al., J. Gen. Virol., 2002, 83, 3075-3084; Harley et al., Clin. Microbiol. Rev., 2001, 14, 909-932). Based on genomic sequence data from six of the seven serocomplexes, alphaviruses have been grouped into three large groups VEE/EEE, SFV and SIN. The VEE-EEE group is exclusively made up of New World viruses with a distribution in North America, South America and Central America. Members of this group include EEE, VEE, EVE, MUC and PIX. The SF group is primarily Old World, but contains one member (MAY) that is found in South America. Other members of the SF group include SF, MID, CHIK, ONN, RR, BF, GET, SAG, BEB and UNA. The SIN group is also primarily Old World, with the exception of AURA, which is a New World virus related to SIN and can be found in Brazil and Argentina. Other members of this group include SIN, WHA, BAB and KYZ. WEE, HJ and FM are considered recombinant viruses and are thus not included in any of the three groups. NDU and Buggy Creek are currently unclassified.

**[0006]** Many members of the alphavirus genus pose a significant health risk to humans, as well as horses, in many different geographic regions. EEE and WEE both cause a fatal encephalitis in humans and horses; however, EEE is more virulent with a mortality rate up to 50%, compared with 3-4% for WEE. VEE can also cause disease in humans and horses, but symptoms are typically flu-like and rarely lead to encephalitis. The geographic distribution for the encephalitis viruses is primarily in the Americas ("The Springer Index of Viruses," pgs. 1148-1155, Tidona and Darai eds., 2001, Springer, New York; Strauss and Strauss, Microbiol. Rev., 1994, 58, 491-562).

**[0007]** The SIN group of Old World viruses, including RR, ONN and CHIK, have been associated with outbreaks of acute and persistent arthritis and arthralgia (joint pain) in humans. Epidemics of acute, debilitating arthralgia have been caused by ONN and CHIK in Africa and Asia. RR, which is the etiological agent of epidemic polyarthritits, is endemic to Australia and caused a major epidemic throughout the Pacific islands in 1979. The outbreak affected over 50,000 people on the island of Fiji. Other alphaviruses have been linked to acute and persistent arthralgia in northern Europe and South Africa. Although each virus induces a somewhat different disease, infection with RR, ONN or CHIK typically causes symptoms such as generalized to severe joint pain, fever, rash, headache, nausea, myalgia and lymphadenitis. It has been reported that arthralgia associated with alphavirus infection can persist for months or years.

CHIK has also been associated with a fatal hemorrhagic condition ("The Springer Index of Viruses," pgs. 1148-1155, Tidona and Darai eds., 2001, Springer, New York; Strauss and Strauss, Microbiol. Rev., 1994, 58, 491-562; Hossain et al., J. Gen. Virol., 2002, 83, 3075-3084).

[0008] Another alphavirus causing human disease and mortality is MAY, which is found in the Caribbean and South America. Mayaro virus infection causes fever, rash and arthropathy (diseases of the joint), and exhibits a mortality rate of up to 7% ("The Springer Index of Viruses," pgs. 1148-1155, Tidona and Darai eds., 2001, Springer, New York).

## **B. Bioagent Detection**

[0009] A problem in determining the cause of a natural infectious outbreak or a bioterrorist attack is the sheer variety of organisms that can cause human disease. There are over 1400 organisms infectious to humans; many of these have the potential to emerge suddenly in a natural epidemic or to be used in a malicious attack by bioterrorists (Taylor et al., Philos. Trans. R. Soc. London B. Biol. Sci., 2001, 356, 983-989). This number does not include numerous strain variants, bioengineered versions, or pathogens that infect plants or animals.

[0010] Much of the new technology being developed for detection of biological weapons incorporates a polymerase chain reaction (PCR) step based upon the use of highly specific primers and probes designed to selectively detect individual pathogenic organisms. Although this approach is appropriate for the most obvious bioterrorist organisms, like smallpox and anthrax, experience has shown that it is very difficult to predict which of hundreds of possible pathogenic organisms might be employed in a terrorist attack. Likewise, naturally emerging human disease that has caused devastating consequence in public health has come from unexpected families of bacteria, viruses, fungi, or protozoa. Plants and animals also have their natural burden of infectious disease agents and there are equally important biosafety and security concerns for agriculture.

[0011] An alternative to single-agent tests is to do broad-range consensus priming of a gene target conserved across groups of bioagents. Broad-range priming has the potential to generate amplification products across entire genera, families, or, as with bacteria, an entire domain of life. This strategy has been successfully employed using consensus 16S ribosomal RNA primers for determining bacterial diversity, both in environmental samples (Schmidt et al., J. Bact., 1991, 173, 4371-4378) and in natural human flora (Kroes et al., Proc Nat Acad Sci (USA), 1999, 96,

14547-14552). The drawback of this approach for unknown bioagent detection and epidemiology is that analysis of the PCR products requires the cloning and sequencing of hundreds to thousands of colonies per sample, which is impractical to perform rapidly or on a large number of samples.

**[0012]** Conservation of sequence is not as universal for viruses, however, large groups of viral species share conserved protein-coding regions, such as regions encoding viral polymerases or helicases. Like bacteria, consensus priming has also been described for detection of several viral families, including coronaviruses (Stephensen et al., *Vir. Res.*, 1999, 60, 181-189), enteroviruses (Oberste et al., *J. Virol.*, 2002, 76, 1244-51); Oberste et al., *J. Clin. Virol.*, 2003, 26, 375-7); Oberste et al., *Virus Res.*, 2003, 91, 241-8), retroid viruses (Mack et al., *Proc. Natl. Acad. Sci. U. S. A.*, 1988, 85, 6977-81); Seifarth et al., *AIDS Res. Hum. Retroviruses*, 2000, 16, 721-729); Donehower et al., *J. Vir. Methods*, 1990, 28, 33-46), and adenoviruses (Echavarria et al., *J. Clin. Micro.*, 1998, 36, 3323-3326). However, as with bacteria, there is no adequate analytical method other than sequencing to identify the viral bioagent present.

**[0013]** In contrast to PCR-based methods, mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign pathogens, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to identify a particular organism.

**[0014]** There is a need for a method for identification of bioagents which is both specific and rapid, and in which no culture or nucleic acid sequencing is required. Disclosed in U.S. Pre-Grant Publication Nos. 2003-0027135, 2003-0082539, 2003-0228571, 2004-0209260, 2004-0219517 and 2004-0180328, and in U.S. Application Serial Nos. 10/660,997, 10/728,486, 10/754,415 and 10/829,826, all of which are commonly owned and incorporated herein by reference in their entirety, are methods for identification of bioagents (any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus) in an unbiased manner by molecular mass and base composition analysis of "bioagent identifying amplicons" which are obtained by amplification of segments of essential and conserved genes which are involved in, for example, translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy

generation, uptake, secretion and the like. Examples of these proteins include, but are not limited to, ribosomal RNAs, ribosomal proteins, DNA and RNA polymerases, RNA-dependent RNA polymerases, RNA capping and methylation enzymes, elongation factors, tRNA synthetases, protein chain initiation factors, heat shock protein groEL, phosphoglycerate kinase, NADH dehydrogenase, DNA ligases, DNA gyrases and DNA topoisomerases, helicases, metabolic enzymes, and the like.

**[0015]** To obtain bioagent identifying amplicons, primers are selected to hybridize to conserved sequence regions which bracket variable sequence regions to yield a segment of nucleic acid which can be amplified and which is amenable to methods of molecular mass analysis. The variable sequence regions provide the variability of molecular mass which is used for bioagent identification. Upon amplification by PCR or other amplification methods with the specifically chosen primers, an amplification product that represents a bioagent identifying amplicon is obtained. The molecular mass of the amplification product, obtained by mass spectrometry for example, provides the means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass of the amplification product or the corresponding base composition (which can be calculated from the molecular mass of the amplification product) is compared with a database of molecular masses or base compositions and a match indicates the identity of the bioagent. Furthermore, the method can be applied to rapid parallel analyses (for example, in a multi-well plate format) the results of which can be employed in a triangulation identification strategy which is amenable to rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent identification.

**[0016]** The result of determination of a previously unknown base composition of a previously unknown bioagent (for example, a newly evolved and heretofore unobserved virus) has downstream utility by providing new bioagent indexing information with which to populate base composition databases. The process of subsequent bioagent identification analyses is thus greatly improved as more base composition data for bioagent identifying amplicons becomes available.

**[0017]** The present invention provides methods of identifying unknown viruses, including viruses of the Togaviridae family and alphavirus genus. Also provided are oligonucleotide primers, compositions and kits containing the oligonucleotide primers, which define alphaviral

identifying amplicons and, upon amplification, produce corresponding amplification products whose molecular masses provide the means to identify alphaviruses at the sub-species level.

## **SUMMARY OF THE INVENTION**

[0018] The present invention provides primers and compositions comprising pairs of primers, and kits containing the same for use in identification of alphaviruses. The primers are designed to produce alphaviral bioagent identifying amplicons of DNA encoding genes essential to alphavirus replication. The invention further provides compositions comprising pairs of primers and kits containing the same, which are designed to provide species and sub-species characterization of alphaviruses.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

[0019] Figure 1 is a process diagram illustrating a representative primer selection process.

[0020] Figure 2 is a representative process diagram for identification and determination of the quantity of a bioagent in a sample.

[0021] Figure 3 is a pseudo four-dimensional plot of expected base compositions of alphavirus identifying amplicons obtained from amplification with primer pair no: 316 the epidemic, epizootic VEEV viruses of classes IAB-IC, ID and IIIA (which have the potential to cause severe disease in humans and animals) can be distinguished from the enzootic VEE types IE, IF, I, IIIB, IIIC, IV, V, and VI, which, in turn, are generally distinguishable from each other.

## **DETAILED DESCRIPTION**

[0022] In the context of the present invention, a “bioagent” is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents include, but are not limited, to cells, including but not limited to human clinical samples, cell cultures, bacterial cells and other pathogens), viruses, viroids, fungi, protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a “pathogen” is a bioagent which causes a disease or disorder.

[0023] As used herein, “intelligent primers” are primers that are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. By the term “highly conserved,” it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity among all or at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of species or strains.

[0024] As used herein, “broad range survey primers” are intelligent primers designed to identify an unknown bioagent at the genus level. In some cases, broad range survey primers are able to identify unknown bioagents at the species or sub-species level. As used herein, “division-wide primers” are intelligent primers designed to identify a bioagent at the species level and “drill-down” primers are intelligent primers designed to identify a bioagent at the sub-species level. As used herein, the “sub-species” level of identification includes, but is not limited to, strains, subtypes, variants, and isolates.

[0025] As used herein, a “bioagent division” is defined as group of bioagents above the species level and includes but is not limited to, orders, families, classes, clades, genera or other such groupings of bioagents above the species level.

[0026] As used herein, a “sub-species characteristic” is a genetic characteristic that provides the means to distinguish two members of the same bioagent species. For example, one viral strain could be distinguished from another viral strain of the same species by possessing a genetic change (e.g., for example, a nucleotide deletion, addition or substitution) in one of the viral genes, such as the RNA-dependent RNA polymerase. In this case, the sub-species characteristic that can be identified using the methods of the present invention, is the genetic change in the viral polymerase.

[0027] As used herein, the term “bioagent identifying amplicon” refers to a polynucleotide that is amplified from a bioagent in an amplification reaction and which 1) provides enough variability to distinguish each individual bioagent and 2) whose molecular mass is amenable to molecular mass determination.



[0028] As used herein, a “base composition” is the exact number of each nucleobase (A, T, C and G) in a given sequence.

[0029] As used herein, a “base composition signature” (BCS) is the exact base composition (i.e., the number of A, T, G and C nucleobases) determined from the molecular mass of a bioagent identifying amplicon.

[0030] As used herein, a “base composition probability cloud” is a representation of the diversity in base composition resulting from a variation in sequence that occurs among different isolates of a given species. The “base composition probability cloud” represents the base composition constraints for each species and is typically visualized using a pseudo four-dimensional plot.

[0031] As used herein, a “wobble base” is a variation in a codon found at the third nucleotide position of a DNA triplet. Variations in conserved regions of sequence are often found at the third nucleotide position due to redundancy in the amino acid code.

[0032] In the context of the present invention, the term “unknown bioagent” may mean either: (i) a bioagent whose existence is known (such as the well known bacterial species *Staphylococcus aureus* for example) but which is not known to be in a sample to be analyzed, or (ii) a bioagent whose existence is not known (for example, the SARS coronavirus was unknown prior to April 2003). For example, if the method for identification of coronaviruses disclosed in commonly owned U.S. Patent Serial No. 10/829,826 (incorporated herein by reference in its entirety) was to be employed prior to April 2003 to identify the SARS coronavirus in a clinical sample, both meanings of “unknown” bioagent are applicable since the SARS coronavirus was unknown to science prior to April, 2003 and since it was not known what bioagent (in this case a coronavirus) was present in the sample. On the other hand, if the method of U.S. Patent Serial No. 10/829,826 was to be employed subsequent to April 2003 to identify the SARS coronavirus in a clinical sample, only the first meaning (i) of “unknown” bioagent would apply since the SARS coronavirus became known to science subsequent to April 2003 and since it was not known what bioagent was present in the sample.

[0033] As used herein, “triangulation identification” means the employment of more than one bioagent identifying amplicons for identification of a bioagent.

[0034] In the context of the present invention, “viral nucleic acid” includes, but is not limited to, DNA, RNA, or DNA that has been obtained from viral RNA, such as, for example, by performing a reverse transcription reaction. Viral RNA can either be single-stranded (of positive or negative polarity) or double-stranded.

[0035] As used herein, the term “etiology” refers to the causes or origins, of diseases or abnormal physiological conditions.

[0036] As used herein, the term “nucleobase” is synonymous with other terms in use in the art including “nucleotide,” “deoxynucleotide,” “nucleotide residue,” “deoxynucleotide residue,” “nucleotide triphosphate (NTP),” or deoxynucleotide triphosphate (dNTP).

[0037] The present invention provides methods for detection and identification of bioagents in an unbiased manner using bioagent identifying amplicons. Intelligent primers are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions to yield a bioagent identifying amplicon which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding base composition signature (BCS) of the amplification product is then matched against a database of molecular masses or base composition signatures. Furthermore, the method can be applied to rapid parallel multiplex analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

[0038] Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination.

[0039] Unlike bacterial genomes, which exhibit conservation of numerous genes (i.e. housekeeping genes) across all organisms, viruses do not share a gene that is essential and conserved among all virus families. Therefore, viral identification is achieved within smaller

groups of related viruses, such as members of a particular virus family or genus. For example, RNA-dependent RNA polymerase is present in all single-stranded RNA viruses and can be used for broad priming as well as resolution within the virus family.

**[0040]** In some embodiments of the present invention, at least one viral nucleic acid segment is amplified in the process of identifying the bioagent. Thus, the nucleic acid segments that can be amplified by the primers disclosed herein and that provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as bioagent identifying amplicons.

**[0041]** In some embodiments of the present invention, bioagent identifying amplicons comprise from about 45 to about 200 nucleobases (i.e. from about 45 to about 200 linked nucleosides). One of ordinary skill in the art will appreciate that the invention embodies compounds of 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, and 200 nucleobases in length, or any range therewithin.

**[0042]** It is the combination of the portions of the bioagent nucleic acid segment to which the primers hybridize (hybridization sites) and the variable region between the primer hybridization sites that comprises the bioagent identifying amplicon.

**[0043]** In some embodiments, bioagent identifying amplicons amenable to molecular mass determination which are produced by the primers described herein are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example. Thus, in some embodiments, bioagent identifying amplicons are larger than 200 nucleobases and are amenable

to molecular mass determination following restriction digestion. Methods of using restriction enzymes and cleavage primers are well known to those with ordinary skill in the art.

[0044] In some embodiments, amplification products corresponding to bioagent identifying amplicons are obtained using the polymerase chain reaction (PCR) which is a routine method to those with ordinary skill in the molecular biology arts. Other amplification methods may be used such as ligase chain reaction (LCR), low-stringency single primer PCR, and multiple strand displacement amplification (MDA) which are also well known to those with ordinary skill.

[0045] Intelligent primers are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. In some embodiments, the highly conserved sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity, or between about 99-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of intelligent primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent. Bioagent identifying amplicons are ideally specific to the identity of the bioagent.

[0046] Identification of bioagents can be accomplished at different levels using intelligent primers suited to resolution of each individual level of identification. Broad range survey intelligent primers are designed with the objective of identifying a bioagent as a member of a particular division (e.g., an order, family, class, clade, genus or other such grouping of bioagents above the species level of bioagents). As a non-limiting example, members of the alphavirus genus may be identified as such by employing broad range survey intelligent primers such as primers which target nsP1 or nsP4. In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species or sub-species level.

[0047] Division-wide intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, eastern equine encephalitis (EEE) virus, western equine encephalitis (WEE) virus and Venezuelan equine encephalitis (VEE) virus can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey

intelligent primers may provide sufficient identification resolution to accomplishing this identification objective.

[0048] Drill-down intelligent primers are designed with the objective of identifying a bioagent at the sub-species level (including strains, subtypes, variants and isolates) based on sub-species characteristics. As one non-limiting example, subtypes IC, ID and IE of Venezuelan equine encephalitis virus can be distinguished from each other using drill-down primers. Drill-down intelligent primers are not always required for identification at the sub-species level because broad range survey intelligent primers may provide sufficient identification resolution to accomplishing this identification objective.

[0049] A representative process flow diagram used for primer selection and validation process is outlined in Figure 1. For each group of organisms, candidate target sequences are identified (200) from which nucleotide alignments are created (210) and analyzed (220). Primers are then designed by selecting appropriate priming regions (230) which then makes possible the selection of candidate primer pairs (240). The primer pairs are then subjected to *in silico* analysis by electronic PCR (ePCR) (300) wherein bioagent identifying amplicons are obtained from sequence databases such as GenBank or other sequence collections (310) and checked for specificity *in silico* (320). Bioagent identifying amplicons obtained from GenBank sequences (310) can also be analyzed by a probability model which predicts the capability of a given amplicon to identify unknown bioagents such that the base compositions of amplicons with favorable probability scores are then stored in a base composition database (325). Alternatively, base compositions of the bioagent identifying amplicons obtained from the primers and GenBank sequences can be directly entered into the base composition database (330). Candidate primer pairs (240) are validated by *in vitro* amplification by a method such as PCR analysis (400) of nucleic acid from a collection of organisms (410). Amplification products thus obtained are analyzed to confirm the sensitivity, specificity and reproducibility of the primers used to obtain the amplification products (420).

[0050] Many of the important pathogens, including the organisms of greatest concern as biological weapons agents, have been completely sequenced. This effort has greatly facilitated the design of primers and probes for the detection of unknown bioagents. The combination of broad-range priming with division-wide and drill-down priming has been used very successfully

in several applications of the technology, including environmental surveillance for biowarfare threat agents and clinical sample analysis for medically important pathogens.

[0051] Synthesis of primers is well known and routine in the art. The primers may be conveniently and routinely made through the well-known technique of solid phase synthesis. Equipment for such synthesis is sold by several vendors including, for example, Applied Biosystems (Foster City, CA). Any other means for such synthesis known in the art may additionally or alternatively be employed.

[0052] The primers are employed as compositions for use in methods for identification of viral bioagents as follows: a primer pair composition is contacted with nucleic acid (such as, for example, DNA from a DNA virus, or DNA reverse transcribed from the RNA of an RNA virus) of an unknown viral bioagent. The nucleic acid is then amplified by a nucleic acid amplification technique, such as PCR for example, to obtain an amplification product that represents a bioagent identifying amplicon. The molecular mass of each strand of the double-stranded amplification product is determined by a molecular mass measurement technique such as mass spectrometry for example, wherein the two strands of the double-stranded amplification product are separated during the ionization process. In some embodiments, the mass spectrometry is electrospray Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) or electrospray time of flight mass spectrometry (ESI-TOF-MS). A list of possible base compositions can be generated for the molecular mass value obtained for each strand and the choice of the correct base composition from the list is facilitated by matching the base composition of one strand with a complementary base composition of the other strand. The molecular mass or base composition thus determined is then compared with a database of molecular masses or base compositions of analogous bioagent identifying amplicons for known viral bioagents. A match between the molecular mass or base composition of the amplification product and the molecular mass or base composition of an analogous bioagent identifying amplicon for a known viral bioagent indicates the identity of the unknown bioagent. In some embodiments, the primer pair used is one of the primer pairs of Table 1. In some embodiments, the method is repeated using a different primer pair to resolve possible ambiguities in the identification process or to improve the confidence level for the identification assignment.

[0053] In some embodiments, a bioagent identifying amplicon may be produced using only a single primer (either the forward or reverse primer of any given primer pair), provided an

appropriate amplification method is chosen, such as, for example, low stringency single primer PCR (LSSP-PCR). Adaptation of this amplification method in order to produce bioagent identifying amplicons can be accomplished by one with ordinary skill in the art without undue experimentation.

**[0054]** In some embodiments, the oligonucleotide primers are broad range survey primers which hybridize to conserved regions of nucleic acid encoding nsP1 of all (or between 80% and 100%, between 85% and 100%, between 90% and 100% or between 95% and 100%) known alphaviruses and produce bioagent identifying amplicons. In some embodiments, the oligonucleotide primers are broad range survey primers which hybridize to conserved regions of nucleic acid encoding nsP4 of all (or between 80% and 100%, between 85% and 100%, between 90% and 100% or between 95% and 100%) known alphaviruses and produce bioagent identifying amplicons. As used herein, the term broad range survey primers refers to primers that bind to nucleic acid encoding genes essential to alphavirus replication (e.g., for example, nsP1 and nsP4) of all (or between 80% and 100%, between 85% and 100%, between 90% and 100% or between 95% and 100%) known species of alphaviruses. In some embodiments, the broad range survey primer pairs comprise oligonucleotides ranging in length from 13-35 nucleobases, each of which have from 70% to 100% sequence identity with primer pair number 966, which corresponds to SEQ ID NOs: 21:66. In some embodiments, the broad range survey primer pairs comprise oligonucleotides ranging in length from 13-35 nucleobases, each of which have from 70% to 100% sequence identity with primer pair number 1131, which corresponds to SEQ ID NOs: 33:78.

**[0055]** In some cases, the molecular mass or base composition of a viral bioagent identifying amplicon defined by a broad range survey primer pair does not provide enough resolution to unambiguously identify a viral bioagent at the species level. These cases benefit from further analysis of one or more viral bioagent identifying amplicons generated from at least one additional broad range survey primer pair or from at least one additional division-wide primer pair. The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as triangulation identification.

**[0056]** In other embodiments, the oligonucleotide primers are division-wide primers which hybridize to nucleic acid encoding genes of species within a genus of viruses. In other embodiments, the oligonucleotide primers are drill-down primers which enable the identification

of sub-species characteristics. Drill down primers provide the functionality of producing bioagent identifying amplicons for drill-down analyses such as strain typing when contacted with nucleic acid under amplification conditions. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of viral infections. In some embodiments, sub-species characteristics are identified using only broad range survey primers and division-wide and drill-down primers are not used.

**[0057]** In some embodiments, the primers used for amplification hybridize to and amplify genomic DNA, DNA of bacterial plasmids, DNA of DNA viruses or DNA reverse transcribed from RNA of an RNA virus.

**[0058]** In some embodiments, the primers used for amplification hybridize directly to viral RNA and act as reverse transcription primers for obtaining DNA from direct amplification of viral RNA. Methods of amplifying RNA using reverse transcriptase are well known to those with ordinary skill in the art and can be routinely established without undue experimentation.

**[0059]** One with ordinary skill in the art of design of amplification primers will recognize that a given primer need not hybridize with 100% complementarity in order to effectively prime the synthesis of a complementary nucleic acid strand in an amplification reaction. Moreover, a primer may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event. (e.g., for example, a loop structure or a hairpin structure). The primers of the present invention may comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity with any of the primers listed in Table 1. Thus, in some embodiments of the present invention, an extent of variation of 70% to 100%, or any range therewithin, of the sequence identity is possible relative to the specific primer sequences disclosed herein. Determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is identical to another 20 nucleobase primer having two non-identical residues has 18 of 20 identical residues ( $18/20 = 0.9$  or 90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of primer 20 nucleobases in length would have  $15/20 = 0.75$  or 75% sequence identity with the 20 nucleobase primer.

**[0060]** Percent homology, sequence identity or complementarity, can be determined by, for example, the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics



Computer Group, University Research Park, Madison WI), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489). In some embodiments, complementarity of primers with respect to the conserved priming regions of viral nucleic acid, is between about 70% and about 80%. In other embodiments, homology, sequence identity or complementarity, is between about 80% and about 90%. In yet other embodiments, homology, sequence identity or complementarity, is at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or is 100%.

**[0061]** In some embodiments, the primers described herein comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or at least 99%, or 100% (or any range therewithin) sequence identity with the primer sequences specifically disclosed herein. Thus, for example, a primer may have between 70% and 100%, between 75% and 100%, between 80% and 100%, and between 95% and 100% sequence identity with SEQ ID NO: 21. Likewise, a primer may have similar sequence identity with any other primer whose nucleotide sequence is disclosed herein.

**[0062]** One with ordinary skill is able to calculate percent sequence identity or percent sequence homology and able to determine, without undue experimentation, the effects of variation of primer sequence identity on the function of the primer in its role in priming synthesis of a complementary strand of nucleic acid for production of an amplification product of a corresponding bioagent identifying amplicon.

**[0063]** In some embodiments of the present invention, the oligonucleotide primers are 13 to 35 nucleobases in length (13 to 35 linked nucleotide residues). These embodiments comprise oligonucleotide primers 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 or 35 nucleobases in length, or any range therewithin.

**[0064]** In some embodiments, any given primer comprises a modification comprising the addition of a non-templated T residue to the 5' end of the primer (i.e., the added T residue does not necessarily hybridize to the nucleic acid being amplified). The addition of a non-templated T residue has an effect of minimizing the addition of non-templated A residues as a result of the non-specific enzyme activity of *Taq* polymerase (Magnuson et al., Biotechniques, 1996, 21, 700-709), an occurrence which may lead to ambiguous results arising from molecular mass analysis.

[0065] In some embodiments of the present invention, primers may contain one or more universal bases. Because any variation (due to codon wobble in the 3<sup>rd</sup> position) in the conserved regions among species is likely to occur in the third position of a DNA (or RNA) triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a “universal nucleobase.” For example, under this “wobble” pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal nucleobases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes et al., *Nucleosides and Nucleotides*, 1995, 14, 1001-1003), the degenerate nucleotides dP or dK (Hill *et al.*), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot et al., *Nucleosides and Nucleotides*, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy- $\beta$ -D-ribofuranosyl)-imidazole-4-carboxamide (Sala et al., *Nucl. Acids Res.*, 1996, 24, 3302-3306).

[0066] In some embodiments, to compensate for the somewhat weaker binding by the wobble base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, 5-propynyluracil which binds to adenine and 5-propynylcytosine and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned and incorporated herein by reference in its entirety. Propynylated primers are described in U.S. Pre-Grant Publication No. 2003-0170682, which is also commonly owned and incorporated herein by reference in its entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

[0067] In some embodiments, to enable broad priming of rapidly evolving RNA viruses, primer hybridization is enhanced using primers and probes containing 5-propynyl deoxy-cytidine and deoxy-thymidine nucleotides. These modified primers and probes offer increased affinity and base pairing selectivity.

[0068] In some embodiments, non-template primer tags are used to increase the melting temperature ( $T_m$ ) of a primer-template duplex in order to improve amplification efficiency. A

non-template tag is at least three consecutive A or T nucleotide residues on a primer which are not complementary to the template. In any given non-template tag, A can be replaced by C or G and T can also be replaced by C or G. Although Watson-Crick hybridization is not expected to occur for a non-template tag relative to the template, the extra hydrogen bond in a G-C pair relative to an A-T pair confers increased stability of the primer-template duplex and improves amplification efficiency for subsequent cycles of amplification when the primers hybridize to strands synthesized in previous cycles.

[0069] In other embodiments, propynylated tags may be used in a manner similar to that of the non-template tag, wherein two or more 5-propynylcytidine or 5-propynyluridine residues replace template matching residues on a primer. In other embodiments, a primer contains a modified internucleoside linkage such as a phosphorothioate linkage, for example.

[0070] In some embodiments, the primers contain mass-modifying tags. Reducing the total number of possible base compositions of a nucleic acid of specific molecular weight provides a means of avoiding a persistent source of ambiguity in determination of base composition of amplification products. Addition of mass-modifying tags to certain nucleobases of a given primer will result in simplification of *de novo* determination of base composition of a given bioagent identifying amplicon from its molecular mass.

[0071] In some embodiments of the present invention, the mass modified nucleobase comprises one or more of the following: for example, 7-deaza-2'-deoxyadenosine-5'-triphosphate, 5-iodo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxycytidine-5'-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'-deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'-triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'-triphosphate, N2-methyl-2'-deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'-triphosphate or thiothymidine-5'-triphosphate. In some embodiments, the mass-modified nucleobase comprises  $^{15}\text{N}$  or  $^{13}\text{C}$  or both  $^{15}\text{N}$  and  $^{13}\text{C}$ .

[0072] In some cases, a molecular mass of a given bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as triangulation identification. Triangulation identification is pursued by analyzing a plurality of

bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen et al., J. Appl. Microbiol., 1999, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

[0073] In some embodiments, the triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR where multiple primers are employed in the same amplification reaction mixture, or PCR in multi-well plate format wherein a different and unique pair of primers is used in multiple wells containing otherwise identical reaction mixtures. Such multiplex and multi-well PCR methods are well known to those with ordinary skill in the arts of rapid throughput amplification of nucleic acids.

[0074] In some embodiments, the molecular mass of a given bioagent identifying amplicon is determined by mass spectrometry. Mass spectrometry has several advantages, not the least of which is high bandwidth characterized by the ability to separate (and isolate) many molecular peaks across a broad range of mass to charge ratio ( $m/z$ ). Thus mass spectrometry is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons.

[0075] In some embodiments, intact molecular ions are generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ESI), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent identifying amplicon. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such

as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

[0076] The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), time of flight (TOF), ion trap, quadrupole, magnetic sector, Q-TOF, and triple quadrupole.

[0077] Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, a base composition signature (BCS) is the exact base composition determined from the molecular mass of a bioagent identifying amplicon. In one embodiment, a BCS provides an index of a specific gene in a specific organism.

[0078] In some embodiments, conversion of molecular mass data to a base composition is useful for certain analyses. As used herein, a base composition is the exact number of each nucleobase (A, T, C and G).

[0079] RNA viruses depend on error-prone polymerases for replication and therefore their nucleotide sequences (and resultant base compositions) drift over time within the functional constraints allowed by selection pressure. Base composition probability distribution of a viral species or group represents a probabilistic distribution of the above variation in the A, C, G and T base composition space and can be derived by analyzing base compositions of all known isolates of that particular species.

[0080] In some embodiments, assignment of base compositions to experimentally determined molecular masses is accomplished using base composition probability clouds. Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building base composition probability clouds around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A pseudo four-dimensional plot can be used to visualize the concept of base composition probability clouds. Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual

bioagents. Areas where clouds overlap indicate regions that may result in a misclassification, a problem which is overcome by a triangulation identification process using bioagent identifying amplicons not affected by overlap of base composition probability clouds.

[0081] In some embodiments, base composition probability clouds provide the means for screening potential primer pairs in order to avoid potential misclassifications of base compositions. In other embodiments, base composition probability clouds provide the means for predicting the identity of a bioagent whose assigned base composition was not previously observed and/or indexed in a bioagent identifying amplicon base composition database due to evolutionary transitions in its nucleic acid sequence. Thus, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition or sequence in order to make the measurement.

[0082] The present invention provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to identify a given bioagent. Furthermore, the process of determination of a previously unknown base composition for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate base composition databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in base composition databases.

[0083] In some embodiments, the identity and quantity of an unknown bioagent can be determined using the process illustrated in Figure 2. Primers (500) and a known quantity of a calibration polynucleotide (505) are added to a sample containing nucleic acid of an unknown bioagent. The total nucleic acid in the sample is then subjected to an amplification reaction (510) to obtain amplification products. The molecular masses of amplification products are determined (515) from which are obtained molecular mass and abundance data. The molecular mass of the bioagent identifying amplicon (520) provides the means for its identification (525) and the molecular mass of the calibration amplicon obtained from the calibration polynucleotide (530) provides the means for its identification (535). The abundance data of the bioagent identifying amplicon is recorded (540) and the abundance data for the calibration data is recorded (545), both of which are used in a calculation (550) which determines the quantity of unknown bioagent in the sample.

[0084] A sample comprising an unknown bioagent is contacted with a pair of primers which provide the means for amplification of nucleic acid from the bioagent, and a known quantity of a polynucleotide that comprises a calibration sequence. The nucleic acids of the bioagent and of the calibration sequence are amplified and the rate of amplification is reasonably assumed to be similar for the nucleic acid of the bioagent and of the calibration sequence. The amplification reaction then produces two amplification products: a bioagent identifying amplicon and a calibration amplicon. The bioagent identifying amplicon and the calibration amplicon should be distinguishable by molecular mass while being amplified at essentially the same rate. Effecting differential molecular masses can be accomplished by choosing as a calibration sequence, a representative bioagent identifying amplicon (from a specific species of bioagent) and performing, for example, a 2-8 nucleobase deletion or insertion within the variable region between the two priming sites. The amplified sample containing the bioagent identifying amplicon and the calibration amplicon is then subjected to molecular mass analysis by mass spectrometry, for example. The resulting molecular mass analysis of the nucleic acid of the bioagent and of the calibration sequence provides molecular mass data and abundance data for the nucleic acid of the bioagent and of the calibration sequence. The molecular mass data obtained for the nucleic acid of the bioagent enables identification of the unknown bioagent and the abundance data enables calculation of the quantity of the bioagent, based on the knowledge of the quantity of calibration polynucleotide contacted with the sample.

[0085] In some embodiments, construction of a standard curve where the amount of calibration polynucleotide spiked into the sample is varied, provides additional resolution and improved confidence for the determination of the quantity of bioagent in the sample. The use of standard curves for analytical determination of molecular quantities is well known to one with ordinary skill and can be performed without undue experimentation.

[0086] In some embodiments, multiplex amplification is performed where multiple bioagent identifying amplicons are amplified with multiple primer pairs which also amplify the corresponding standard calibration sequences. In this or other embodiments, the standard calibration sequences are optionally included within a single vector which functions as the calibration polynucleotide. Multiplex amplification methods are well known to those with ordinary skill and can be performed without undue experimentation.

[0087] In some embodiments, the calibrant polynucleotide is used as an internal positive control to confirm that amplification conditions and subsequent analysis steps are successful in producing a measurable amplicon. Even in the absence of copies of the genome of a bioagent, the calibration polynucleotide should give rise to a calibration amplicon. Failure to produce a measurable calibration amplicon indicates a failure of amplification or subsequent analysis step such as amplicon purification or molecular mass determination. Reaching a conclusion that such failures have occurred is in itself, a useful event.

[0088] In some embodiments, the calibration sequence is comprised of DNA. In some embodiments, the calibration sequence is comprised of RNA.

[0089] In some embodiments, the calibration sequence is inserted into a vector which then itself functions as the calibration polynucleotide. In some embodiments, more than one calibration sequence is inserted into the vector that functions as the calibration polynucleotide. Such a calibration polynucleotide is herein termed a "combination calibration polynucleotide." The process of inserting polynucleotides into vectors is routine to those skilled in the art and can be accomplished without undue experimentation. Thus, it should be recognized that the calibration method should not be limited to the embodiments described herein. The calibration method can be applied for determination of the quantity of any bioagent identifying amplicon when an appropriate standard calibrant polynucleotide sequence is designed and used. The process of choosing an appropriate vector for insertion of a calibrant is also a routine operation that can be accomplished by one with ordinary skill without undue experimentation.

[0090] Bioagents that can be identified by the methods of the present invention include RNA viruses. The genomes of RNA viruses can be positive-sense single-stranded RNA, negative-sense single-stranded RNA or double-stranded RNA. Examples of RNA viruses with positive-sense single-stranded genomes include, but are not limited to members of the Caliciviridae, Picornaviridae, Flaviviridae, Togaviridae, Retroviridae and Coronaviridae families. Examples of RNA viruses with negative-sense single-stranded RNA genomes include, but are not limited to, members of the Filoviridae, Rhabdoviridae, Bunyaviridae, Orthomyxoviridae, Paramyxoviridae and Arenaviridae families. Examples of RNA viruses with double-stranded RNA genomes include, but are not limited to, members of the Reoviridae and Birnaviridae families.



[0091] In some embodiments of the present invention, RNA viruses are identified by first obtaining RNA from an RNA virus, or a sample containing or suspected of containing an RNA virus, obtaining corresponding DNA from the RNA by reverse transcription, amplifying the DNA to obtain one or more amplification products using one or more pairs of oligonucleotide primers that bind to conserved regions of the RNA viral genome, which flank a variable region of the genome, determining the molecular mass or base composition of the one or more amplification products and comparing the molecular masses or base compositions with calculated or experimentally determined molecular masses or base compositions of known RNA viruses, wherein at least one match identifies the RNA virus. Methods of isolating RNA from RNA viruses and/or samples containing RNA viruses, and reverse transcribing RNA to DNA are well known to those of skill in the art.

[0092] Alphaviruses represent RNA virus examples of bioagents which can be identified by the methods of the present invention. Alphaviruses are extremely diverse at the nucleotide and protein sequence levels and are thus difficult to detect and identify using currently available diagnostic techniques.

[0093] In one embodiment of the present invention, the alphavirus target gene is nsP4, which is the viral RNA-dependent RNA polymerase. In another embodiment, the target gene is nsP1, which functions to cap and methylate the 5' end of genomic and subgenomic alphaviral RNAs.

[0094] In other embodiments of the present invention, the intelligent primers produce bioagent identifying amplicons within stable and highly conserved regions of alphaviral genomes. The advantage to characterization of an amplicon in a highly conserved region is that there is a low probability that the region will evolve past the point of primer recognition, in which case, the amplification step would fail. Such a primer set is thus useful as a broad range survey-type primer. In another embodiment of the present invention, the intelligent primers produce bioagent identifying amplicons in a region which evolves more quickly than the stable region described above. The advantage of characterization bioagent identifying amplicon corresponding to an evolving genomic region is that it is useful for distinguishing emerging strain variants.

[0095] The present invention also has significant advantages as a platform for identification of diseases caused by emerging viruses. The present invention eliminates the need for prior knowledge of bioagent sequence to generate hybridization probes. Thus, in another embodiment,

the present invention provides a means of determining the etiology of a virus infection when the process of identification of viruses is carried out in a clinical setting and, even when the virus is a new species never observed before. This is possible because the methods are not confounded by naturally occurring evolutionary variations (a major concern for characterization of viruses which evolve rapidly) occurring in the sequence acting as the template for production of the bioagent identifying amplicon. Measurement of molecular mass and determination of base composition is accomplished in an unbiased manner without sequence prejudice.

[0096] Another embodiment of the present invention also provides a means of tracking the spread of any species or strain of virus when a plurality of samples obtained from different locations are analyzed by the methods described above in an epidemiological setting. In one embodiment, a plurality of samples from a plurality of different locations are analyzed with primers which produce bioagent identifying amplicons, a subset of which contain a specific virus. The corresponding locations of the members of the virus-containing subset indicate the spread of the specific virus to the corresponding locations.

[0097] The present invention also provides kits for carrying out the methods described herein. In some embodiments, the kit may comprise a sufficient quantity of one or more primer pairs to perform an amplification reaction on a target polynucleotide from a bioagent to form a bioagent identifying amplicon. In some embodiments, the kit may comprise from one to fifty primer pairs, from one to twenty primer pairs, from one to ten primer pairs, or from two to five primer pairs. In some embodiments, the kit may comprise one or more primer pairs recited in Table 1.

[0098] In some embodiments, the kit may comprise one or more broad range survey primer(s), division wide primer(s), or drill-down primer(s), or any combination thereof. A kit may be designed so as to comprise particular primer pairs for identification of a particular bioagent. For example, a broad range survey primer kit may be used initially to identify an unknown bioagent as a member of the alphavirus genus. Another example of a division-wide kit may be used to distinguish eastern equine encephalitis virus, western equine encephalitis virus and Venezuelan equine encephalitis virus from each other. A drill-down kit may be used, for example, to distinguish different subtypes of Venezuelan equine encephalitis virus, or to identify genetically engineered alphaviruses. In some embodiments, any of these kits may be combined to comprise a combination of broad range survey primers and division-wide primers so as to be able to identify the species of an unknown bioagent.

[0099] In some embodiments, the kit may contain standardized calibration polynucleotides for use as internal amplification calibrants. Internal calibrants are described in commonly owned U.S. Patent Application Serial No: 60/545,425 which is incorporated herein by reference in its entirety.

[0100] In some embodiments, the kit may also comprise a sufficient quantity of reverse transcriptase (if an RNA virus is to be identified for example), a DNA polymerase, suitable nucleoside triphosphates (including any of those described above), a DNA ligase, and/or reaction buffer, or any combination thereof, for the amplification processes described above. A kit may further include instructions pertinent for the particular embodiment of the kit, such instructions describing the primer pairs and amplification conditions for operation of the method. A kit may also comprise amplification reaction containers such as microcentrifuge tubes and the like. A kit may also comprise reagents or other materials for isolating bioagent nucleic acid or bioagent identifying amplicons from amplification, including, for example, detergents, solvents, or ion exchange resins which may be linked to magnetic beads. A kit may also comprise a table of measured or calculated molecular masses and/or base compositions of bioagents using the primer pairs of the kit.

[0101] Any of the primers, primer pairs, or compositions described herein can be used in the preparation of a kit for diagnosing or detecting the presence or absence of an alphavirus in any sample.

[0102] While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same. In order that the invention disclosed herein may be more efficiently understood, examples are provided below. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting the invention in any manner.

## EXAMPLES

**[0103] Example 1: Selection of Primers That Define Alphavirus Identifying Amplicons**

[0104] For design of primers that define alphaviral bioagent identifying amplicons, relevant sequences from, for example, GenBank were obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 45 to about 200 nucleotides in length and

distinguish species and/or sub-species from each other by their molecular masses or base compositions. A typical process shown in Figure 1 is employed.

[0105] A database of expected base compositions for each primer region is generated using an *in silico* PCR search algorithm, such as (ePCR). An existing RNA structure search algorithm (Macke et al., Nucl. Acids Res., 2001, 29, 4724-4735, which is incorporated herein by reference in its entirety) has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (SantaLucia, Proc. Natl. Acad. Sci. U.S.A., 1998, 95, 1460-1465, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs.

[0106] Table 1 represents a collection of primers (sorted by forward primer name) designed to identify alphaviruses using the methods described herein. Primer sites were identified on two essential alphaviral genes, nsP1 (the RNA capping and methylation enzyme) and nsP4, the RNA-dependent RNA polymerase). The forward or reverse primer name shown in Table 1 indicates the gene region of the viral genome to which the primer hybridizes relative to a reference sequence. For example, the forward primer name AV\_NC001449\_888\_901P\_F indicates a forward primer that hybridizes to residues 888-901 of an alphavirus reference sequence represented by GenBank Accession No. NC001449 (SEQ ID NO: 1). In Table 1, U<sup>a</sup> = 5-propynyluracil; C<sup>a</sup> = 5-propynylcytosine; \* = phosphorothioate linkage. The primer pair number is an in-house database index number.

**Table 1: Primer Pairs for Identification of Alphavirus Bioagents**

Primer pair number	For. primer name	Forward sequence	For. SEQ ID NO:	Reverse primer name	Reverse sequence	Reverse SEQ ID NO:
302	AV_NC001449_159_178P_F	AATGCTAGAGC <sup>a</sup> GU <sup>a</sup> U <sup>a</sup> TU <sup>a</sup> C <sup>a</sup> GCA	2	AV_NC001449_225_244P_R	GCACTTC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup> CAGGAT	47
303	AV_NC001449_159_178P_F	AATGCTAGAGC <sup>a</sup> GU <sup>a</sup> U <sup>a</sup> TU <sup>a</sup> C <sup>a</sup> GCA	3	AV_NC001449_225_244P_2_R	GCACTTC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup> TAGGAT	48
304	AV_NC001449_162_178P_F	GCTAGAGC <sup>a</sup> GU <sup>a</sup> U <sup>a</sup> TU <sup>a</sup> C <sup>a</sup> GC A	4	AV_NC001449_231_247P_R	GGCGC <sup>a</sup> AC <sup>a</sup> U <sup>a</sup> TC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup>	49
306	AV_NC001449_888_904P_F	TGC <sup>a</sup> GAAGGGU <sup>a</sup> ACGTCGT	5	AV_NC001449_972_991P_R	TTGC <sup>a</sup> AGCACAAGAATC <sup>a</sup> C <sup>a</sup> CTC	50
307	AV_NC001449_1057_1072P_F	TGU <sup>a</sup> GTGAC <sup>a</sup> C <sup>a</sup> AGAU <sup>a</sup> GAC	6	AV_NC001449_1122_1135P_R	TGGU <sup>a</sup> U <sup>a</sup> GAGC <sup>a</sup> C <sup>a</sup> C <sup>a</sup> AAC	51
314	AV_NC001449_159_178P_F	AATGCTAGAGCGTTTTCGC A	7	AV_NC001449_225_244_R	GCACTTCCAATGTCCAG GAT	52
315	AV_NC001449_159_178P_F	AATGCTAGAGCGTTTTCGC A	8	AV_NC001449_225_244_2_R	GCACTTCCAATGTCTAG GAT	53
316	AV_NC0014	GCTAGAGCGTTTTCGCA	9	AV_NC001449	GGCGCACTTCCAATGTC	54

Primer pair number	For. primer name	Forward sequence	For. SEQ ID NO:	Reverse primer name	Reverse sequence	Reverse SEQ ID NO:
	49_162_17 8_F			_231_247_R		
317	AV_NC0014 49_888_90 1_F	TGCGAAGGGTACGT	10	AV_NC001449 _972_991_R	TTGCAGCACAAGAATCC CTC	55
318	AV_NC0014 49_888_90 4_F	TGCGAAGGGTACGTCGT	11	AV_NC001449 _972_991_R	TTGCAGCACAAGAATCC CTC	56
319	AV_NC0014 49_1057_1 072_F	TGTGTGACCAGATGAC	12	AV_NC001449 _1122_1135_ R	TGGTTGAGCCCAAC	57
494	AV_NC0014 49_158_17 8P_F	TAATGCTAGAGC <sup>a</sup> GU <sup>a</sup> TU <sup>a</sup> <sup>a</sup> C <sup>a</sup> GCA	13	AV_NC001449 _225_245P_R	TGCACTTC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup> CAGGAT	58
494	AV_NC0014 49_158_17 8P_F	TAATGCTAGAGC <sup>a</sup> GU <sup>a</sup> TU <sup>a</sup> <sup>a</sup> C <sup>a</sup> GCA	14	AV_NC001449 _225_245P_R	TGCACTTC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup> CAGGAT	59
495	AV_NC0014 49_158_17 8P_F	TAATGCTAGAGC <sup>a</sup> GU <sup>a</sup> TU <sup>a</sup> <sup>a</sup> C <sup>a</sup> GCA	15	AV_NC001449 _225_245P_2 R	TGCACTTC <sup>a</sup> C <sup>a</sup> AAU <sup>a</sup> GU <sup>a</sup> C <sup>a</sup> TAGGAT	60
496	AV_NC0014 49_161_17 8P_F	TGCTAGAGC <sup>a</sup> GU <sup>a</sup> TU <sup>a</sup> C <sup>a</sup> G CA	16	AV_NC001449 _231_248P_R	TGGCGC <sup>a</sup> AC <sup>a</sup> U <sup>a</sup> TC <sup>a</sup> C <sup>a</sup> AA U <sup>a</sup> GU <sup>a</sup> C	61
497	AV_NC0014 49_887_90 1P_F	TTGC <sup>a</sup> GAAGGGU <sup>a</sup> ACGT	17	AV_NC001449 _972_992P_R	TTTGC <sup>a</sup> AGCACAAGAAT C <sup>a</sup> C <sup>a</sup> CTC	62
498	AV_NC0014 49_887_90 4P_F	TTGC <sup>a</sup> GAAGGGU <sup>a</sup> ACGTCGT	18	AV_NC001449 _972_992P_R	TTTGC <sup>a</sup> AGCACAAGAAT C <sup>a</sup> C <sup>a</sup> CTC	63
498	AV_NC0014 49_887_90 4P_F	TTGC <sup>a</sup> GAAGGGU <sup>a</sup> ACGTCGT	19	AV_NC001449 _972_992P_R	TTTGC <sup>a</sup> AGCACAAGAAT C <sup>a</sup> C <sup>a</sup> CTC	64
499	AV_NC0014 49_1056_1 072P_F	TTGU <sup>a</sup> GTGAC <sup>a</sup> C <sup>a</sup> AGAU <sup>a</sup> GA C	20	AV_NC001449 _1122_1136P R	TTGGU <sup>a</sup> U <sup>a</sup> GAGC <sup>a</sup> C <sup>a</sup> C <sup>a</sup> AA C	65
966	AV_NC_001 449_151_1 78_F	TCCATGCTAATGCTAGAGC GTTTTTCGCA	21	AV_NC_00144 9_225_248_R	TGGCGCACTTCCAATGT CCAGGAT	66
967	AV_NC_001 449_881_9 04_F	TGTCAGTTGCGAAGGGTAC GTCGT	22	AV_NC_00144 9_972_1000_ R	TCTGTCACTTTGCAGCA CAAGAATCCCTC	67
968	AV_NC_001 449_158_1 78_F	TAATGCTAGAGCGTTTTTCG CA	23	AV_NC_00144 9_225_245_R	TGCACTTCCAATGTCCA GGAT	68
969	AV_NC_001 449_887_9 04_F	TTGCGAAGGGTACGTCGT	24	AV_NC_00144 9_972_992_R	TTTGCAGCACAAGAATC CCTC	69
970	AV_NC_001 449_156_1 78P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> AATGCTAGAGCGTT TTCGCA	25	AV_NC_00144 9_225_247P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> GCACTTCCAATG TCCAGGAT	70
971	AV_NC_001 449_885_9 04P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> TGCGAAGGGTACGT CGT	26	AV_NC_00144 9_972_994P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> TTGCAGCACAAG AATCCCTC	71
972	AV_NC_001 449_155_1 78P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> AATGCTAGAGCG TTTTTCGCA	27	AV_NC_00144 9_225_248P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> GCCTTCCAA TGTCAGGAT	72
973	AV_NC_001 449_884_9 04P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> TGCGAAGGGTAC GTCGT	28	AV_NC_00144 9_972_995P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> TTGCAGCACA AGAATCCCTC	73
974	AV_NC_001 449_154_1 78P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> AATGCTAGAGC GTTTTTCGCA	29	AV_NC_00144 9_225_249P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> GCCTTCCAA ATGTCCAGGAT	74
975	AV_NC_001 449_883_9 04P_F	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> TGCGAAGGGTA CGTCGT	30	AV_NC_00144 9_972_996P_ R	U <sup>a</sup> C <sup>a</sup> C <sup>a</sup> U <sup>a</sup> TTGCAGCAC AAGAATCCCTC	75
976	AV_NC_001 449_153_1 78_F	TCCTTCAATGCTAGAGCGT TTTCGCA	31	AV_NC_00144 9_225_250_R	TCCTTCACTTCCAAT GTCCAGGAT	76
977	AV_NC_001 449_882_9 04_F	TCCTTCTGCGAAGGGTACG TCGT	32	AV_NC_00144 9_972_997_R	TCCTTCTGAGCACA GAATCCCTC	77

Primer pair number	For. primer name	Forward sequence	For. SEQ ID NO:	Reverse primer name	Reverse sequence	Reverse SEQ ID NO:
1131	AV_NC_001 449_1045_ 1072_F	TGCCAGCTACACTGTGCGA CCAGATGAC	33	AV_NC_00144 9_1122_1149 R	TGACGACTATCCGCTGG TTGAGCCCAAC	78
1146	AV_NC_001 449_878_9 01_F	TATTGTTCAGTTGCGAAGGG TACGT	34	AV_NC_00144 9_972_998_R	TGTCACCTTTGCAACACA AGAATCCCTC	79
1147	AV_NC_001 449_878_9 01_2_F	TATTGTTCAGTTGCGACGGG TACGT	35	AV_NC_00144 9_972_998_R	TGTCACCTTTGCAACACA AGAATCCCTC	80
1148	AV_NC_001 449_876_9 01_F	TCTATAGTCAGTTGCGACG GGTACGT	36	AV_NC_00144 9_972_998_2 R	TGTCACCTTTGCGACACA AGAATCCCTC	81
1149	AV_NC_001 449_1045_ 1075_F	TGTCAGCTACATTGTGTGA CCAAATGACTGG	37	AV_NC_00144 9_1122_1149 2_R	TGACGACTATCCGCTGG TTGAGCCCAAC	82
1150	AV_NC_001 449_1045_ 1075_2_F	TACCAGCCACACTTTGCGA TCAGATGACAGG	38	AV_NC_00144 9_1122_1149 2_R	TGACGACTATCCGCTGG TTGAGCCCAAC	83
2048	AV_NC_001 449_151_1 78_2_F	TCCATGCTAACGCCAGAGC GTTTTCGCA	39	AV_NC_00144 9_225_251_R	TGCTGGTGCACCTTCCAA TATCCAGGAT	84
2049	AV_NC_001 449_151_1 78_2_F	TCCATGCTAACGCCAGAGC GTTTTCGCA	40	AV_NC_00144 9_228_251_R	TGCCCGTGCCTGCCTA TGTCCTCAA	85
2050	AV_NC_001 449_62_86 F	TGACGTAGACCCCGAGAGT CCGTTT	41	AV_NC_00144 9_147_171_R	TCGCTCTGGCATTAGCA TGGTCATT	86
2051	AV_NC_001 449_6971_ 6997_F	TGGCGCTATGATGAAATCT GGAATGTT	42	AV_NC_00144 9_7083_7106 R	TATGTTGTTCGTCGCCGA TGAACGC	87
2052	AV_NC_001 449_6971_ 6997_F	TGGCGCTATGATGAAATCT GGAATGTT	43	AV_NC_00144 9_7086_7109 R	TACGATGTTGTCTGTCGC CGATGAA	88
2053	AV_NC_001 449_7082_ 7105_F	TGCCTTCATCGGCGATGAC AACAT	44	AV_NC_00144 9_7134_7158 R	TCCAAGTGGCGCACCTG TCTGCCAT	89
2054	AV_NC_001 449_6742_ 6772_F	TGTCGGCCGAGGATTTTGA TGCTATCATAGC	45	AV_NC_00144 9_6816_6841 R	TCATCTTGGCTTTTGTCTC AAAGGAGGC	90
2055	AV_NC_001 449_6254_ 6280_F	TGCGGTACCGTCACCATTT CAGAACAC	46	AV_NC_00144 9_6318_6347 R	TGGTAGTTCTCTCATTT GTGTGACGTTGCA	91

### [0107] Example 2: One-step RT-PCR of RNA Virus Samples

[0108] RNA was isolated from virus-containing samples according to methods well known in the art. To generate bioagent identifying amplicons for RNA viruses, a one-step RT-PCR protocol was developed. All RT-PCR reactions were assembled in 50 µl reactions in the 96 well microtiter plate format using a Packard MPII liquid handling robotic platform and MJ Dyad<sup>®</sup> thermocyclers (MJ research, Waltham, MA). The RT-PCR reaction consisted of 4 units of Amplitaq Gold<sup>®</sup>, 1.5x buffer II (Applied Biosystems, Foster City, CA), 1.5 mM MgCl<sub>2</sub>, 0.4 M betaine, 10 mM DTT, 20 mM sorbitol, 50 ng random primers (Invitrogen, Carlsbad, CA), 1.2 units Supersasin (Ambion, Austin, TX), 100 ng polyA DNA, 2 units Superscript III (Invitrogen, Carlsbad, CA), 400 ng T4 Gene 32 Protein (Roche Applied Science, Indianapolis, IN), 800 µM dNTP mix, and 250 nM of each primer.

[0109] The following RT-PCR conditions were used to amplify the sequences used for mass spectrometry analysis: 60°C for 5 minutes, 4°C for 10 minutes, 55°C for 45 minutes, 95°C for 10 minutes followed by 8 cycles of 95 °C for 30 seconds, 48 °C for 30 seconds, and 72 °C for 30 seconds, with the 48 °C annealing temperature increased 0.9 °C after each cycle. The PCR reaction was then continued for 37 additional cycles of 95 °C for 15 seconds, 56 °C for 20 seconds, and 72 °C for 20 seconds. The reaction concluded with 2 minutes at 72 °C.

**[0110] Example 3: Solution Capture Purification of PCR Products for Mass Spectrometry with Ion Exchange Resin-Magnetic Beads**

[0111] For solution capture of nucleic acids with ion exchange resin linked to magnetic beads, 25 µl of a 2.5 mg/mL suspension of BioClon amine terminated supraparamagnetic beads were added to 25 to 50 µl of a PCR (or RT-PCR) reaction containing approximately 10 pM of a typical PCR amplification product. The above suspension was mixed for approximately 5 minutes by vortexing or pipetting, after which the liquid was removed after using a magnetic separator. The beads containing bound PCR amplification product were then washed 3x with 50mM ammonium bicarbonate/50% MeOH or 100mM ammonium bicarbonate/50% MeOH, followed by three more washes with 50% MeOH. The bound PCR amplicon was eluted with 25mM piperidine, 25mM imidazole, 35% MeOH, plus peptide calibration standards.

**[0112] Example 4: Mass Spectrometry and Base Composition Analysis**

[0113] The ESI-FTICR mass spectrometer is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer that employs an actively shielded 7 Tesla superconducting magnet. The active shielding constrains the majority of the fringing magnetic field from the superconducting magnet to a relatively small volume. Thus, components that might be adversely affected by stray magnetic fields, such as CRT monitors, robotic components, and other electronics, can operate in close proximity to the FTICR spectrometer. All aspects of pulse sequence control and data acquisition were performed on a 600 MHz Pentium II data station running Bruker's Xmass software under Windows NT 4.0 operating system. Sample aliquots, typically 15 µl, were extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR data station. Samples were injected directly into a 10 µl sample loop integrated with a fluidics handling system that supplies the 100 µl/hr flow rate to the ESI source. Ions were formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned approximately 1.5 cm from the

metalized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary was biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N<sub>2</sub> was employed to assist in the desolvation process. Ions were accumulated in an external ion reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection into the trapped ion cell where they were mass analyzed. Ionization duty cycles > 99% were achieved by simultaneously accumulating ions in the external ion reservoir during ion detection. Each detection event consisted of 1M data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans were co-added for a total data acquisition time of 74 s.

[0114] The ESI-TOF mass spectrometer is based on a Bruker Daltonics MicroTOF™. Ions from the ESI source undergo orthogonal ion extraction and are focused in a reflectron prior to detection. The TOF and FTICR are equipped with the same automated sample handling and fluidics described above. Ions are formed in the standard MicroTOF™ ESI source that is equipped with the same off-axis sprayer and glass capillary as the FTICR ESI source. Consequently, source conditions were the same as those described above. External ion accumulation was also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF was comprised of 75,000 data points digitized over 75 μs.

[0115] The sample delivery scheme allows sample aliquots to be rapidly injected into the electrospray source at high flow rate and subsequently be electrosprayed at a much lower flow rate for improved ESI sensitivity. Prior to injecting a sample, a bolus of buffer was injected at a high flow rate to rinse the transfer line and spray needle to avoid sample contamination/carryover. Following the rinse step, the autosampler injected the next sample and the flow rate was switched to low flow. Following a brief equilibration delay, data acquisition commenced. As spectra were co-added, the autosampler continued rinsing the syringe and picking up buffer to rinse the injector and sample transfer line. In general, two syringe rinses and one injector rinse were required to minimize sample carryover. During a routine screening protocol a new sample mixture was injected every 106 seconds. More recently a fast wash station for the syringe needle has been implemented which, when combined with shorter acquisition times, facilitates the acquisition of mass spectra at a rate of just under one spectrum/minute.

[0116] Raw mass spectra were post-calibrated with an internal mass standard and deconvoluted to monoisotopic molecular masses. Unambiguous base compositions were derived from the



exact mass measurements of the complementary single-stranded oligonucleotides. Quantitative results are obtained by comparing the peak heights with an internal PCR calibration standard present in every PCR well at 500 molecules per well. Calibration methods are commonly owned and disclosed in U.S. Provisional Patent Application Serial No. 60/545,425.

**[0117] Example 5: *De Novo* Determination of Base Composition of Amplification Products using Molecular Mass Modified Deoxynucleotide Triphosphates**

**[0118]** Because the molecular masses of the four natural nucleobases have a relatively narrow molecular mass range (A = 313.058, G = 329.052, C = 289.046, T = 304.046 – See Table 2), a persistent source of ambiguity in assignment of base composition can occur as follows: two nucleic acid strands having different base composition may have a difference of about 1 Da when the base composition difference between the two strands is G ↔ A (-15.994) combined with C ↔ T (+15.000). For example, one 99-mer nucleic acid strand having a base composition of A<sub>27</sub>G<sub>30</sub>C<sub>21</sub>T<sub>21</sub> has a theoretical molecular mass of 30779.058 while another 99-mer nucleic acid strand having a base composition of A<sub>26</sub>G<sub>31</sub>C<sub>22</sub>T<sub>20</sub> has a theoretical molecular mass of 30780.052. A 1 Da difference in molecular mass may be within the experimental error of a molecular mass measurement and thus, the relatively narrow molecular mass range of the four natural nucleobases imposes an uncertainty factor.

**[0119]** The present invention provides for a means for removing this theoretical 1 Da uncertainty factor through amplification of a nucleic acid with one mass-tagged nucleobase and three natural nucleobases. The term “nucleobase” as used herein is synonymous with other terms in use in the art including “nucleotide,” “deoxynucleotide,” “nucleotide residue,” “deoxynucleotide residue,” “nucleotide triphosphate (NTP),” or deoxynucleotide triphosphate (dNTP).

**[0120]** Addition of significant mass to one of the 4 nucleobases (dNTPs) in an amplification reaction, or in the primers themselves, will result in a significant difference in mass of the resulting amplification product (significantly greater than 1 Da) arising from ambiguities arising from the G ↔ A combined with C ↔ T event (Table 2). Thus, the same the G ↔ A (-15.994) event combined with 5-Iodo-C ↔ T (-110.900) event would result in a molecular mass difference of 126.894. If the molecular mass of the base composition A<sub>27</sub>G<sub>30</sub>**5-Iodo-C**<sub>21</sub>T<sub>21</sub> (33422.958) is compared with A<sub>26</sub>G<sub>31</sub>**5-Iodo-C**<sub>22</sub>T<sub>20</sub>, (33549.852) the theoretical molecular mass difference is +126.894. The experimental error of a molecular mass measurement is not significant with regard to this molecular mass difference. Furthermore, the only base

composition consistent with a measured molecular mass of the 99-mer nucleic acid is  $A_{27}G_{30}5\text{-Iodo-C}_{21}T_{21}$ . In contrast, the analogous amplification without the mass tag has 18 possible base compositions.

**Table 2: Molecular Masses of Natural Nucleobases and the Mass-Modified Nucleobase 5-Iodo-C and Molecular Mass Differences Resulting from Transitions**

Nucleobase	Molecular Mass	Transition	$\Delta$ Molecular Mass
A	313.058	A $\rightarrow$ T	-9.012
A	313.058	A $\rightarrow$ C	-24.012
A	313.058	A $\rightarrow$ 5-Iodo-C	101.888
A	313.058	A $\rightarrow$ G	15.994
T	304.046	T $\rightarrow$ A	9.012
T	304.046	T $\rightarrow$ C	-15.000
T	304.046	T $\rightarrow$ 5-Iodo-C	110.900
T	304.046	T $\rightarrow$ G	25.006
C	289.046	C $\rightarrow$ A	24.012
C	289.046	C $\rightarrow$ T	15.000
C	289.046	C $\rightarrow$ G	40.006
5-Iodo-C	414.946	5-Iodo-C $\rightarrow$ A	-101.888
5-Iodo-C	414.946	5-Iodo-C $\rightarrow$ T	-110.900
5-Iodo-C	414.946	5-Iodo-C $\rightarrow$ G	-85.894
G	329.052	G $\rightarrow$ A	-15.994
G	329.052	G $\rightarrow$ T	-25.006
G	329.052	G $\rightarrow$ C	-40.006
G	329.052	G $\rightarrow$ 5-Iodo-C	85.894

#### [0121] Example 6: Data Processing

[0122] Mass spectra of bioagent identifying amplicons are analyzed independently using a maximum-likelihood processor, such as is widely used in radar signal processing. This processor, referred to as GenX, first makes maximum likelihood estimates of the input to the mass spectrometer for each primer by running matched filters for each base composition aggregate on the input data. This includes the GenX response to a calibrant for each primer.

[0123] The algorithm emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of *a priori* expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database is used to define the mass base count matched filters. The database contains the sequences of known bacterial bioagents and includes threat organisms as well as benign background organisms. The latter is used to estimate and subtract the spectral signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to

form signatures which are then subtracted. The maximum likelihood process is applied to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

[0124] The amplitudes of all base compositions of bioagent identifying amplicons for each primer are calibrated and a final maximum likelihood amplitude estimate per organism is made based upon the multiple single primer estimates. Models of all system noise are factored into this two-stage maximum likelihood calculation. The processor reports the number of molecules of each base composition contained in the spectra. The quantity of amplification product corresponding to the appropriate primer set is reported as well as the quantities of primers remaining upon completion of the amplification reaction.

**[0125] Example 7: Alignment of Alphavirus Sequences using an nsP1 Primer Pair**

[0126] A total of 42 alphavirus sequences, including two strains of EEEV, 20 strains of VEEV, one strain of Chikungunya virus, one strain of Igbo Ora virus, two strains of O'nyong-nyong virus, one strain of Ross River virus, one strain of Sagiyama virus, one strain of Mayaro virus, one strain of Barmah forest virus, two strains of Semliki forest virus, one strain of aura virus, one strain of Ockelbo virus, and seven strains of Sindbois virus were aligned and evaluated for identification of useful priming regions. In a representative example, with reference to the reference sequence NC\_001449 (SEQ ID NO: 1) representing the genome of Venezuelan equine encephalitis virus (VEEV), a pair of primers (no. 316 – SEQ ID NOs: 9:54) was designed to produce an alphavirus identifying amplicon 86 nucleobases long corresponding to positions 162-247 of the nsP1 gene of VEEV. This pair of primers is expected to produce an alphavirus identifying amplicon that can provide the means to identify the virus strains described above.

[0127] As shown in Figure 3, in a pseudo four-dimensional plot of expected base compositions of alphavirus identifying amplicons arising from amplification with primer pair no: 316 the epidemic, epizootic VEEV viruses of classes IAB-IC, ID and IIIA (which have the potential to cause severe disease in humans and animals) can be distinguished from the enzootic VEE types IE, IF, I, IIIB, IIIC, IV, V, and VI, which, in turn, are generally distinguishable from each other.

[0128] Table 3 lists the results of base composition analysis of nine laboratory test isolates of alphaviruses obtained according to the methods described herein by amplification with primer pair 316 to obtain alphavirus identifying amplicons.

**Table 3: Expected and Observed Base Compositions of Alphavirus Identifying Amplicons  
Produced with Primer Pair No: 316 (SEQ ID NOs: 9:54)**

Virus	Strain	Sequence Available	Expected Base Composition [A G C T]	Observed Base Composition [A G C T]
VEE	3908 (subtype IC, 1995)	Yes	[21 23 23 19]	[21 23 23 19]
VEE	66637 (subtype ID, 1981)	Yes	[21 23 23 19]	[21 23 23 19]
VEE	68U201 (Subtype 1E, 1968)	Yes	[22 25 19 20]	[22 25 19 20]
VEE	243937 (subtype 1C, 1992)	Yes	[21 23 23 19]	[21 23 23 19]
WEE	OR71 (71V1658)	Yes	[22 26 19 19]	[22 26 19 19]
WEE	SD83 (R43738)	No	-	[22 26 19 19]
WEE	ON41 (McMillan)	No	-	[22 27 18 19]
WEE	Fleming (Fleming)	No	-	[22 25 19 20]
EEE	(Parker Strain)	Yes	[23 25 19 19]	[23 25 19 19]

**[0129] Example 8: Identification of Six Alphavirus Strains**

**[0130]** Two primers pairs (numbers 966 and 1131) which each amplify a sequence of the alphavirus gene nsP1 were tested for their ability to detect and differentiate among eight different known alphavirus strains using the methods described herein. The strains included in the study were the North American strain of Eastern equine encephalitis virus and the Tonate CaAn 410d, 78V3531, AG80-663, Cabassou CaAr 508 and Everglades Fe3-7c strains of Venezuelan equine encephalitis virus. RT-PCR reactions were spiked with either 10-fold or 100-fold dilutions of virus stock and performed according to the method described in Example 2. Each reaction also contained 500 RNA copies of a calibration sequence to quantitate the amount of virus present in each reaction. The calibration sequence is contained within a combination calibration polynucleotide designated RT-PCR calibrant pVIR001 (SEQ ID NO: 92). This calibration sequence was designed with reference to Venezuelan equine encephalitis virus (VEE) strain 3908, subtype IC (GenBank gi number 20800454) such that all primers disclosed herein with the exception of primer pair numbers 2050-2055, hybridize to the calibration sequence and produce alphavirus calibration amplicons that are distinguishable from alphavirus identifying amplicons. Mass spectral analysis of the alphavirus bioagent identifying amplicons resulted in the correct identification of all six alphavirus strains.

**[0131] Example 9: Identification of Related Alphavirus Species**

**[0132]** A series of eight strains of alphaviruses whose alphavirus identifying amplicon sequences (from primer pairs 966 and 1131) are unknown were analyzed using primer pairs 966 and 1131 by the methods described herein. These experiments were carried out without the presence of a

calibrant. A representative set of results is shown in Table 4 where it is indicated that the "unknown" alphavirus strains can be assigned to related "known" strains.

**Table 4: Representative Result Set of Identification of Alphaviruses with Primer Pair Nos: 966 (SEQ ID NOs: 21:66) and 1131 (SEQ ID NOs: 33:78)**

Sample	Spiked Virus	Primer Pair No:	Base Composition [A G C T]	Match Type	Alphavirus Strain Matched
1	Sindbis Virus	966	[24 25 26 23]	exact	Sindbis virus (NoStrain_14_1, genome strain)
1	Sindbis Virus	1131	[29 26 27 23]	exact	Sindbis virus (DI-2, NoStrain_14_1, genome strain)
2	Nduma Virus	966	[26 27 22 23]	exact	Eastern equine encephalitis virus (North American)
2	Nduma Virus	1131	ND		
3	Middleburg Virus	966	[26 27 22 23]	exact	Eastern equine encephalitis virus (North American)
3	Middleburg Virus	1131	[28 29 27 20]		Deconvolved BC (none)
4	Mayaro Virus	966	[31 24 22 21]		Mayaro virus (NoStrain_5_1, NoStrain_6_2)
4	Mayaro Virus	1131	[26 30 26 23]	mass adjust +/- 1	Venezuelan equine encephalitis virus (78V3531)
5	Highlands J Virus	966	[28 28 22 20]	no match	Deconvolved BC (none)
5	Highlands J Virus	1131	[28 31 28 18]	cloud offset [-1 0 0 1]	Venezuelan equine encephalitis virus (243937, 3908, 6119, 66457, 66637, 71-180; 600035-71-180/4, 83U434, P676, PMCHO5, SH3, TC-83, Trinidad donkey, V198, ZPC738)
6	Getah Virus	966	[25 24 26 23]	cloud offset [-1 0 0 1]	Sindbis virus (NoStrain_14_1, genome strain)
6	Getah Virus	1131	ND		
1	Barmah Virus	966	[30 23 23 22]	exact	Barmah Forest virus (BH2193)
1	Barmah Virus	1131	[25 27 30 22]	no match	Deconvolved BC (none)
2	Semliki Virus	966	[28 23 26 21]	exact	Semliki forest virus (A7-74, DI-19, DI-6, Defective RNA particle, L10, genome strain)
2	Semliki	1131	ND		

	Virus				
--	-------	--	--	--	--

[0133] Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference (including, but not limited to, journal articles, U.S. and non-U.S. patents, patent application publications, international patent application publications, gene bank accession numbers, internet web sites, and the like) cited in the present application is incorporated herein by reference in its entirety. Those skilled in the art will appreciate that numerous changes and modifications may be made to the embodiments of the invention and that such changes and modifications may be made without departing from the spirit of the invention. It is therefore intended that the appended claims cover all such equivalent variations as fall within the true spirit and scope of the invention.

**WHAT IS CLAIMED IS:**

1. An oligonucleotide primer up to 35 nucleobases in length comprising at least 70% sequence identity with SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, or 46.
2. An oligonucleotide primer up to 35 nucleobases in length comprising at least 70% sequence identity with SEQ ID NO: 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, or 91.
3. A composition comprising a primer of claim 1 or claim 2.
4. A composition comprising at least one oligonucleotide primer pair, each primer of the pair comprising up to 35 nucleobases in length, and each primer of the pair comprising at least 70% sequence identity with a primer of claim 1 or claim 2.
5. The composition of claim 4 wherein the at least one oligonucleotide primer pair is SEQ ID NOs: 2:47, 3:48, 4:49, 5:50, 6:51, 7:52, 8:53, 9:54, 10:55, 11:56, 12:57, 13:58, 14:59, 15:60, 16:61, 17:62, 18:63, 19:64, 20:65, 21:66, 22:67, 23:68, 24:69, 25:70, 26:71, 27:72, 28:73, 29:74, 30:75, 31:76, 32:77, 33:78, 34:79, 35:80, 36:81, 37:82, 38:83, 39:84, 40:85, 41:86, 42:87, 43:88, 44:89, 45:90, or 46:91.
6. The composition of claim 4 wherein either or both of the oligonucleotide primers comprises at least one modified nucleobase.
7. The composition of claim 4 wherein either or both of the oligonucleotide primers comprises a non-templated T residue on the 5'-end.
8. The composition of claim 4 wherein either or both of the oligonucleotide primers comprises at least one non-template tag.
9. The composition of claim 4 wherein either or both of the oligonucleotide primers comprises at least one molecular mass modifying tag.

10. A kit comprising the composition of claim 4.
11. The kit of claim 10 further comprising at least one calibration polynucleotide.
12. The kit of claim 10 or claim 11 comprising at least one ion exchange resin linked to magnetic beads.
13. A method for identification of an unknown alphavirus comprising:
  - amplifying nucleic acid from said alphavirus using the composition of any one of claims 4-9 to obtain an amplification product;
  - measuring the molecular mass of said amplification product;
  - optionally, determining the base composition of said amplification product from said molecular mass; and
  - comparing said molecular mass or base composition with a plurality of molecular masses or base compositions of known alphaviral bioagent identifying amplicons, wherein a match between said molecular mass or base composition and a member of said plurality of molecular masses or base compositions identifies said unknown alphavirus.
14. The method of claim 13 wherein said molecular mass is measured by mass spectrometry.
15. A method of determining the presence or absence of an alphavirus species in a sample comprising:
  - amplifying nucleic acid from said sample using the composition of any one of claims 4-9 to obtain an amplification product;
  - determining the molecular mass of said amplification product;
  - optionally, determining the base composition of said amplification product from said molecular mass; and
  - comparing said molecular mass or base composition of said amplification product with the known molecular masses or base compositions of one or more known alphavirus species bioagent identifying amplicons, wherein a match between said molecular mass or base composition of said amplification product and the molecular mass or base composition of one or more known alphavirus species bioagent identifying amplicons indicates the presence of said alphavirus species in said sample.



16. The method of claim 15 wherein said molecular mass is measured by mass spectrometry.

17. A method for determination of the quantity of an unknown alphavirus in a sample comprising:

contacting said sample with the composition of claim 4 and a known quantity of a calibration polynucleotide comprising a calibration sequence;

concurrently amplifying nucleic acid from said alphavirus in said sample with the composition of any one of claims 4-9 and amplifying nucleic acid from said calibration polynucleotide in said sample with the same composition of any one of claims 4-9 to obtain a first amplification product comprising an alphaviral bioagent identifying amplicon and a second amplification product comprising a calibration amplicon;

determining the molecular mass and abundance for said alphaviral bioagent identifying amplicon and said calibration amplicon; and

distinguishing said alphaviral bioagent identifying amplicon from said calibration amplicon based on molecular mass, wherein comparison of alphaviral bioagent identifying amplicon abundance and calibration amplicon abundance indicates the quantity of alphavirus in said sample.

18. The method of claim 17 further comprising determining the base composition of said alphaviral bioagent identifying amplicon.

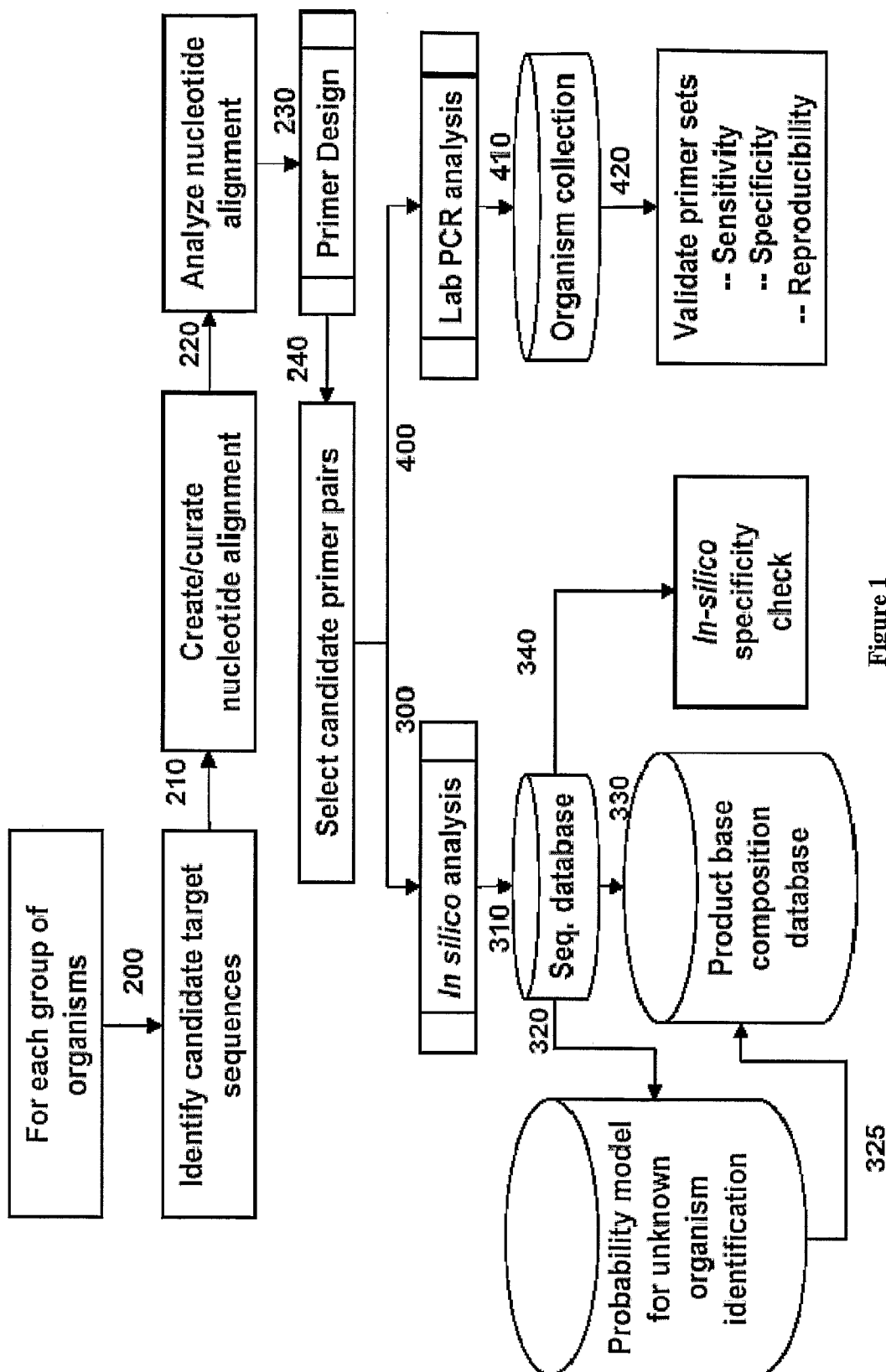


Figure 1

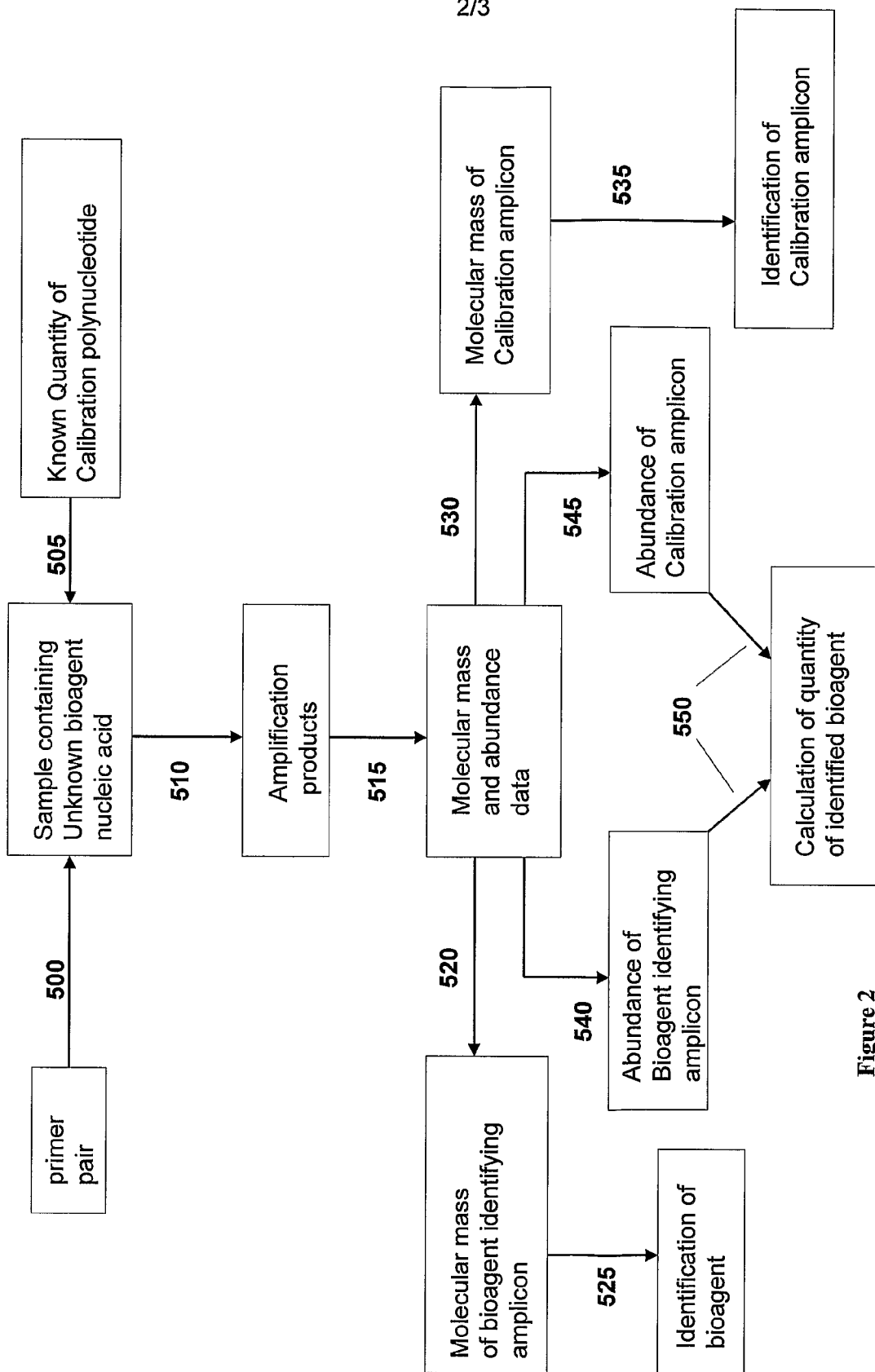


Figure 2

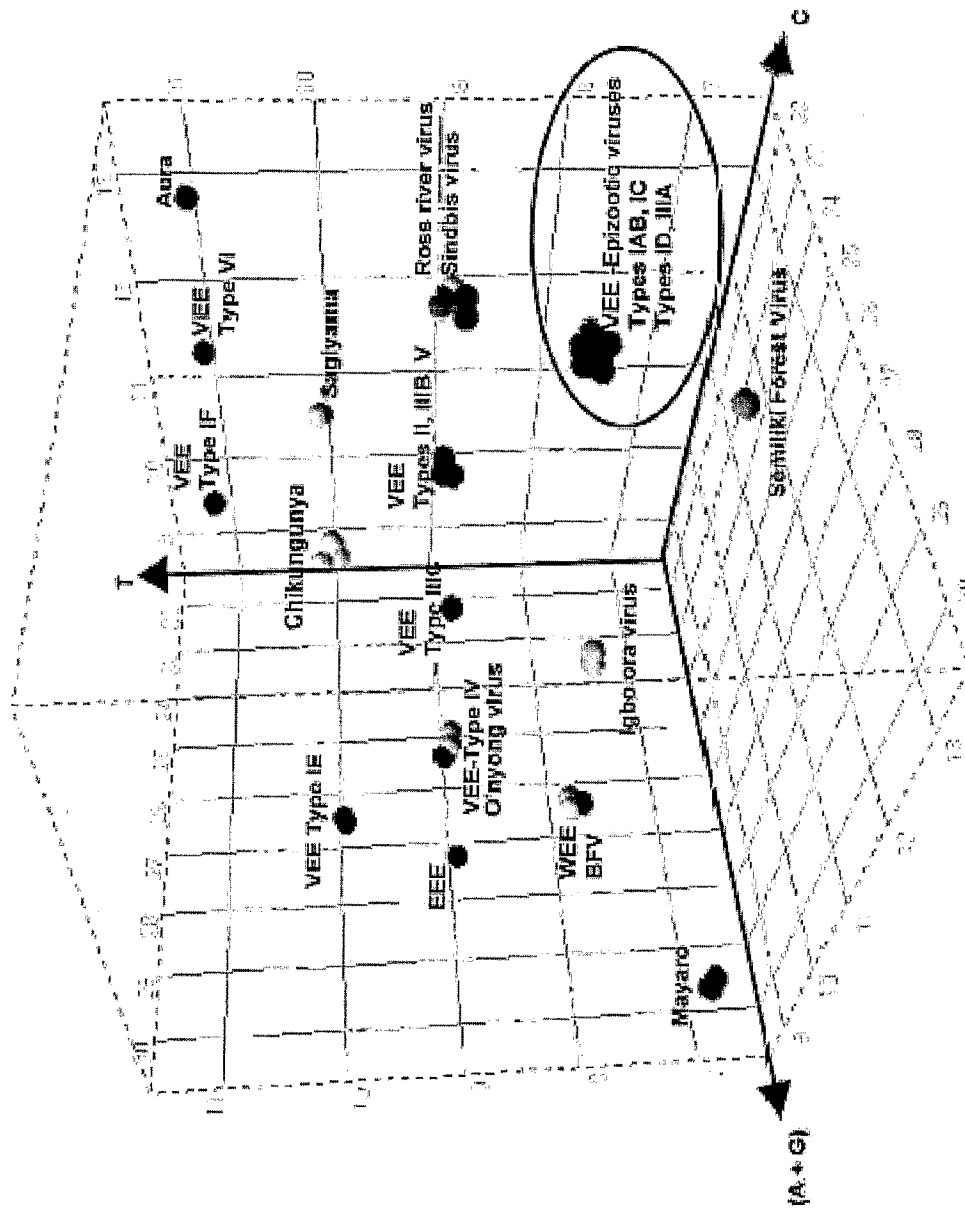


Figure 3

## SEQUENCE LISTING

<110> Isis Pharmaceuticals, Inc.  
Ranjaraajan Sampath  
Thomas A. Hall  
Mark W. Eshoo

<120> COMPOSITIONS FOR USE IN IDENTIFICATION OF ALPHAVIRUSES

<130> IBIS0070-500WO (DIBIS-0054WO)

<150> 60/550,023

<151> 2004-03-03

<160> 92

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 11444

<212> DNA

<213> Venezuelan equine encephalitis virus

<400> 1

atgggcggcg	caagagagaa	gcccaaacca	attacctacc	caaaatggag	aaagttcacg	60
ttgacatcga	ggaagacagc	ccattcctca	gagctttaca	acggagcttc	ccgcagtttg	120
aggtagaagc	caagcaggtc	actgataatg	accatgctaa	tgccagagcg	ttttcgcatc	180
tggcttcaaa	actgatcgaa	acggaggtgg	acccatccga	cacgatcctt	gacattggaa	240
gtgcgcccgc	cgcgagaatg	tattctaagc	ataagtatca	ttgcactgtg	ccgatgagat	300
gtgcggaaga	tccggacaga	ttgtacaagt	atgcaactaa	gctgaagaaa	aattgcaagg	360
aaataactga	caaggaattg	gacaagaaaa	tgaaggagct	cgcgcgcgtc	atgagcgacc	420
ctgacctgga	aactgagact	atgtgcctcc	acgacgatga	gtcatgtcgc	tacgaggggc	480
aagtgcgtgt	ttaccaggat	gtatacgcag	ttgacggacc	gacaagtctc	tatcaccaag	540
ccaacaaggg	agttagagtc	gcctactgga	taggctttga	caccacccct	tttatgttta	600
agaacttggc	tggagcatat	ccatcatact	ctaccaactg	ggccgacgaa	accgtgttaa	660
cggctcgtaa	cataggccta	tgcagctccg	acgtcatgga	gcggtcacgt	agagggatgt	720
ccattcttag	gaagaagtat	ttgaaaccat	ccaataatgt	cctattctct	gttggctcga	780
ccatctacca	cgagaagagg	gacttactga	ggagctggca	cctgocgtct	gtatttcaat	840
tacgtggcaa	gcaaaattac	acatgtcggg	gtgagactat	agttagttag	gacgggtacg	900
tcgttaaaag	aatagctatc	agtcagggcc	tgtatgggaa	gccttcaggc	tatgctgcta	960
cgatgcaccg	cgagggattc	ttgtgctgca	aagtgcacaga	cacattgaac	ggggagaggg	1020
tctcttttcc	cgtgtgcacg	tatgtgccag	ctacattgtg	tgaccaaagt	actggcatac	1080
tggcaacaga	tgtcagtgcg	gacgacgcgc	aaaaactgct	ggttgggctc	aaccagcgca	1140
tagtgcgtcaa	cggtcgcacc	caaagaaaca	ccaataccat	gaagaattat	cttttgcccg	1200
tagtgggcca	ggcattttgct	aggtgggcaa	aggaatataa	ggaagatcaa	gaagatgaga	1260
ggccactagg	actacgagat	agacagttag	tcattggggtg	ctgctgggct	tttagaaggc	1320
acaagataac	atctattttat	aagcgcccag	atacccaaac	catcatcaaa	gtgaacagcg	1380
atttcactc	attcgtgctg	cccaggatag	gcagtaacac	actggagatc	gggctgagaa	1440
cgagaatcag	gaaaatgcta	gaagagcaca	aggagccgtc	acctotcatt	actgccgagg	1500
acatacaaga	ggctaagtgc	gcagccgatg	aggctaagga	agtgcgtgaa	gccgaggagc	1560
tgcgcgctgc	tctaccacct	ttggcagctg	attttgagga	gcccactctg	gaagccgatg	1620
tcgacttgat	gttacaagag	gctggggccg	gctcagtggg	gacacctcgt	ggcttgataa	1680
aggttaccag	ctatgccggc	gaggacaaga	tgggtctctt	cgcagtgtct	tctccacagg	1740
ctgtactcaa	gagtgaagaa	ctatcttgca	ttcacctctc	cgtgaacaa	gtcatagtga	1800
taacacactc	tggccgaaaa	ggcggttatg	cgtggaacc	ctaccatgga	aaagtagtgg	1860
tgcagagggg	acatgcaata	ccgtccagg	actttcaagc	tctgagtga	agtgcaccca	1920
tcgtgtacaa	cgaacgagag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag	1980
gagcgctgaa	cacagatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg	2040
aatacctgta	cgacatcgac	aggaaacaat	gcgtcaagaa	agaattagtc	actgggctag	2100
ggcttacagg	cgagctgggtg	gatcctccct	tccatgaatt	tgctacgag	agtctgagaa	2160
cacgtccggc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccggggtcag	2220

gcaagtctgg	catcattaaa	agcgcagtc	ccaaaaaaga	tctgggtggg	agcgccaaga	2280
aagaaaaactg	cgcagaaata	ataagggaag	tcaagaaaaat	gaaagggctg	gacgtcaatg	2340
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaaaaca	ccccgtagag	accctgtata	2400
ttgacgaagc	ttttgcttgt	catgcaggca	ctctcagagc	gctcatagcc	atcataagac	2460
ctaaaaaggc	agtgtctctgc	ggggatccaa	aacagtggtg	ctttttcaat	atgatgtgcc	2520
tgaagtgca	ttttaaccac	gagatttgca	cgcaggtctt	ccacaaaagc	atctctcgcc	2580
gttgcaactaa	atocgtgact	tccgtcgtct	caaccttgtt	ttaogacaaa	aggatgagaa	2640
cgacgaaccc	gaaagagact	aagatttgta	ttgacactac	tggcagtacc	aaaccgaagc	2700
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca	2760
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggtc	gacccgtaaa	ggcgtgtatg	2820
ccgttcggta	caaggtgaat	gaaaatcccc	tgtacgcacc	cacctcagaa	catgtgaacg	2880
tcctactgac	ccgcacggag	gaccgtatcg	tgtggaaaac	actagccggt	gatccatgga	2940
taaaaatact	gacggccaag	tatcctggga	acttcactgc	cacgatagag	gaatggcaag	3000
cagagcatga	tgccatcatg	aggcacatct	tggagagacc	ggaccctacc	gacgttttcc	3060
aaaataaggc	gaacgttgtt	tgggccaagg	cttttggtgcc	ggtactgaag	actgcaggca	3120
tagacatgac	cactgaacaa	tggaaactct	tggattactt	cgaaacggac	aaagctcact	3180
cagcagagat	agtattgaac	caactatgcg	tgaggttctt	tggactcgac	ctggactccg	3240
gtctattttc	tgcacccact	gttcggttat	ccattaggaa	taatcactgg	gataattccc	3300
cgtcgccctaa	catgtacggg	ttgaataaag	aagtgggtccg	ccagctctcc	cgcaggtacc	3360
cacaactgcc	tcgagcagtt	gccaccggaa	gagtcctatga	catgaacact	ggcacgtctgc	3420
gcaattatga	tcgcgcgata	aatctagtac	ctgtgaacag	aagactgcct	catgcttttag	3480
tcctccacca	taatgaacac	ccacagagtg	actttttctt	attcgtcagc	aaactgaagg	3540
gcagaactgt	cttgggtggc	ggggagaagt	tgtccgtccc	aggcaaaaag	gtcgactggg	3600
tgtcagacca	gcctgaggct	acctttagag	ctcggctgga	tttaggtatc	ccaggtgacg	3660
tgccaaaata	cgacattgta	tttattaaag	tgaggactcc	atataaatac	catcattatc	3720
agcagtgta	agaccacgcc	attaagctta	gtatggtgac	caagaaagct	tgtctgcatt	3780
tgaatcccgg	cggaaacctgc	gtcagcatag	gttatggta	cgtgacagg	gccagcgaga	3840
gcactattgg	tgcatacgcg	cggcagttca	agttctcccg	ggtatgcaaa	ccgaaatcct	3900
cacatgaagta	gacagaagta	ctgtttgtat	tcattgggta	cgatcgcaag	gcccgtacgc	3960
acaatcctta	caagctttca	tctacettga	ccaacatcta	tacaggttcc	agactccacg	4020
aagccggatg	cgcacccctca	tatcatgtgg	tgcgagggga	tattgccacg	gccaccgaag	4080
gagtgatcat	aaatgctgct	aacagcaaag	gacaacctgg	cggaggggtg	tgcggagcgc	4140
tgtataagaa	attcccggaa	agcttcgatt	tacagccgat	cgaagtagga	aaagcgcgac	4200
tgggtcaaagg	tgcagctaaa	catatcatte	atgccgtagg	acaaaacttc	aacaaaagttt	4260
cgggaagtga	aggggacaaa	cagttggcag	aggcttatga	gtccatcgct	aaaattgtca	4320
acgataacaa	ttacaagtca	gtacgcatte	cactgttgct	cacccgcctc	ttttccggga	4380
acaaagatcg	actaacccaa	tcattgaacc	atgtgtgac	agcttttagac	accactgatg	4440
cagatgtagc	catatactgc	agggacaaga	aatgggaaat	gactctcaag	gaagcagtg	4500
ctaggagaga	agcagtggag	gagatatgca	tatcagacga	ctcttcgggtg	acagaaccgg	4560
atgcagagct	ggtgagggta	catccgaaga	gttctttggc	tggaaaggaag	ggctacagca	4620
caagtgatgg	caagactttc	tcataatttg	aagggaccaa	atttcaccag	gcggccaagg	4680
atatagcaga	aattaatgcc	atgtggccag	ttgcaacgga	ggccaatgag	caagtatgca	4740
tgtatatcct	cggtgaaagc	atgagcagca	ttaggtcgaa	atgccccgtc	gaggagtccg	4800
aagcctccac	accacctagc	acgctgcctt	gcttgtgcat	ccatgctatg	actccagaaa	4860
gagtacaacg	cctaaaagcc	tcacgtccag	aacaaattac	tgtgtgctca	tcctttccat	4920
tgccgaagta	tagaatcact	ggtgtgcaga	agatccagtg	ctcccagcct	atactgttct	4980
caccgaaggt	gcctgcgtac	attcatccac	ggaagtacct	cgtggaaaaca	ccaccggtag	5040
aagagactcc	ggagtgcgcg	gcagagaacc	aatccacaga	ggggacacct	gaacaaccag	5100
cacttgtaaa	cgtggatgca	accaggacta	gaatgcctga	accgatcatc	attgaagagg	5160
aagaagagga	tagtataagt	ttgctgtcag	acggcccggc	ccaccagggtg	ctgcaagtcg	5220
aggcagacat	tcacgggtcg	ccttctgtat	ccagctcatc	ctgggtccatt	cctcatgcat	5280
ccgactttga	tgtggacagc	ttatccatcc	ttgacacctt	ggatggagct	agcgtgacca	5340
gcggggagct	gtcagccgag	actaactcct	acttcgcaag	gagcatggag	tttcggggcg	5400
gaccggtgcc	tgcgcctcga	accgtattca	ggaaccttcc	acatcccgca	ccgcgcacaa	5460
gaacaccgcc	acttgacacac	agcagggcc	gctcgagaac	tagcctagtt	tcaccccgcc	5520
caggcgtgaa	tagggtgatt	actagagagg	agctcgaggc	gcttaccocg	tcocgcgctc	5580
ctagcaggtc	ggcctcaaga	actagcctgg	tctctaacc	gccaggcgta	aatagggtga	5640
ttacaagaga	ggagtttgag	gcgttcgtag	cacaacaaca	atgacgggtt	gacgcgggtg	5700
catacatctt	ttcctocgat	accggtcaag	ggcattttaca	acaaaaatca	gtaaggcaaa	5760
cgggtgttatc	cgaagtgggtg	ttggagagga	ccgaattgga	gatttcgtat	gccccgcgcc	5820
tcgaccagga	aaaagaagaa	ctactacgca	agaaattaca	gctgaatccc	acacctgcta	5880

acagaagcag	ataccagtc	aggagggtg	agaatatgaa	agccataaca	gctagacgta	5940
ttctgcaagg	cctagggcat	tatttgaagg	cagaaggaaa	agtggagtgc	tatcgaaccc	6000
tgcatcctgt	tcctttgtat	tcattctagt	tgaatcgtgc	tttttcaagc	cccaaggtcg	6060
cagtgggaagc	ctgcaatgcc	atgctgaaag	aaaattttcc	gactgtagct	tcctactgta	6120
ttattccaga	gtacgatgcc	tattctggaca	tggttgacgg	cgtttcttgt	tgcttagaca	6180
ctgccagttt	ttgccctgcg	aagctgcgca	gctttccaaa	gaaacactcc	tattttggaac	6240
ccacaatacg	gtcggcagtg	ccatcagcga	ttcagaacac	gctccagaac	gtcctggcag	6300
ctgccacaaa	aagaaattgc	aacgtcacgc	aatgagaga	attgcccgta	ttggattcgg	6360
ctgcctttta	tgtggaatgc	ttcaagaaat	atgctgcaa	taatgaatat	tggaagacgt	6420
ttaaagaaaa	ccccatcagg	cttactgaag	aaaattgtgt	aaattacatt	actaaattaa	6480
aaggacaaaa	agctgctgct	ctttttgcga	agacacataa	tttgaatatg	ttacaggaca	6540
taccaatgga	cagggtttgta	atggacttaa	agagggacgt	gaaagtgact	ccaggaaaca	6600
aacatactga	agaacggccc	aaggtagagg	tgattcaggc	tgccgatcca	ctagcgacag	6660
cggatctgtg	cggaaatccac	cgggagttgg	ttaggagatt	aaatgctgtc	ctgcttccga	6720
acatccatac	actgtttgac	atgtcggctg	aagactttga	cgtattattt	gccgagcatt	6780
tccagcctgg	ggactgtgta	ctggaactg	acattgcgtc	gtttgataaa	agtgaggacg	6840
acgccatggc	tctgaccgcg	ttaatgatcc	tggaagacct	aggagtggac	gcagagctgt	6900
tgacgtgat	tgaggcggct	ttcggcgaaa	tatcatcaat	acatttgccc	acaaaaacta	6960
aatttaaat	oggagocatt	atgaaatccg	gaatgttcct	cacactgttt	gtgaacacag	7020
tcattcaaat	cgtaatcgca	agcagagcgt	taagagagcg	gctaaccgga	tcaccatgtg	7080
cagcattcat	tggagatgac	aatatcgtga	aaggagtcaa	atctgacaaa	ttaatggcag	7140
acaggtgcgc	cacttggttg	aacatggaag	tcaagatcat	agacgcctg	gtgggcgaga	7200
aagcgcccta	tttttgtgga	gggtttatct	tgtgtgactc	cgtgaccggc	acagcgtgcc	7260
gtgtggcaga	ccccctaaaa	aggctgttta	agcttgga	acccctggca	gtagacgatg	7320
aacatgacga	tgacaggaga	agggcattac	acgaagagtc	aacacgctgg	aatcgagtgg	7380
gaattcttcc	agagctgtgt	aaggcagtag	aatcaaggta	tgaaaccgta	ggaacttcca	7440
tcatagttaa	ggccatgact	actctagcta	gcagtgttaa	atcattcagc	tacctgagag	7500
gggcccctat	aactctctac	ggctaacctg	aatggactac	gacatagtct	agtcgcgcaa	7560
gatgtttccg	ttccaaccaa	tgtatccgat	cgaccgaatg	ccctatcgta	acccgttcgc	7620
ggccccgcgc	aggccctggt	ttcccagaac	cgaccctttt	ctggcgatgc	aggtgcagga	7680
attaaccgcg	tcgatggcta	acctgacgtt	caagcaacgc	cgggacgcgc	cacctgaggg	7740
gccacctgct	aagaaaccta	agagggaggc	cccgcaaaag	caaaaagggg	gaggccaagg	7800
gaagaagaag	aagaaccagg	ggaagaagaa	ggccaagacg	gggcccgccta	atccgaaggc	7860
acagagtgga	aacaagaaga	agcccaacaa	gaaaccaggc	aagagacagc	gcatggtcat	7920
gaaattggaa	tctgacaaga	cattcccaat	tatgctggaa	gggaagatta	acggctacgc	7980
ttgctgtgtc	ggagggaagt	tattcaggcc	gatgcacgtg	gaaggcaaga	tcgacaacga	8040
cgttctggcc	gcacttaaga	cgaagaaagc	atccaaatat	gatcttgagt	atgcagatgt	8100
gccacagaac	atgcgggccc	gtacacccat	gagaagcccc	aaggctatta	aaggctatta	8160
cagctggcat	catggagcag	ttcaatatga	aatgggcgt	ttcacgggtc	caaaaggagt	8220
tggggccaa	ggagacagcg	gaagacccat	tctggataat	cagggaacgg	tggtcgctat	8280
tgtgctggga	ggtgtgaatg	aaggatctag	gacagccctt	tcagtogtca	tgtggaacga	8340
gaaggagata	actgtgaagt	atactccgga	gaactgcgag	caatggtcac	tagtgaccac	8400
tatgtgctg	ctcgccaatg	tgacgttccc	atgtgcccga	ccaccaattt	gctacgacag	8460
aaaaccagca	gagacttttg	ccatgctcag	cgttaacgtt	gacaaccogg	gctacgatga	8520
gctgctggaa	gcagctgtta	agtgcgccgg	aagaaaaagg	agatctaccg	aggagctgtt	8580
taaggagtat	aagctaacgc	gcccttacat	ggccagatgc	atcagatgtg	ccgttgggag	8640
ctgccatagt	ccaatagcaa	ttgaggcagt	gaagagcgac	gggcacgacg	gctatgttag	8700
acttcagact	tcctcgcagt	atggcctgga	ttcctctggc	aaactaaagg	gaaggactat	8760
gcggtatgat	atgcacggga	ccattgaaga	gataccacta	catcaagtgt	cactccacac	8820
atctcgcccc	tgtcacattg	tggtatggga	tggttatatt	ctgcttgcta	ggtgcccggc	8880
aggggactcc	atcaccatgg	aatttaagaa	aggttcagtc	acacactcct	gctcagtgcc	8940
gtatgaagtg	aaattttaatc	ctgtaggcag	agaactctac	actcatccac	cagaacacgg	9000
agcagagcaa	gcgtgccaa	tctacgcgca	cgatgcacag	aacagaggag	cttatgtcga	9060
gactgcacct	cgggctcag	aagtggacag	cagtttgatt	tccttgagcg	gcagttcagt	9120
caccgtgaca	cctcctgtcg	ggactagcgc	cttggtgaaa	tgcaagtgcg	gcggcacaaa	9180
gatctccgaa	accatcaaca	aggcaaaaca	gttcagccag	tgcaaaaaga	aggagcagtg	9240
cagagcatat	cgactgcaga	atgacaagtg	ggtgtataat	tctgacaaac	tgcccaaaag	9300
agcgggagcc	accctaaaag	gaaaactaca	cgtcccgttc	ttgctggcag	acggcaaatg	9360
caccgtgcct	ctagcaccgg	aacctatgat	aaccttcggt	ttcogatcag	tgtcactgaa	9420
actgcacct	agaatccca	catatctgac	cactcgccaa	cttgctgatg	agcctcatta	9480
cacgcacgag	ctcatatctg	aaccagctgt	taggaatttt	accgtcactg	aaaaggggtg	9540

```

ggagtttgtg tggggaaacc atccgccgaa aaggttttgg gcacaggaaa cagcaccogg 9600
aaatccacat gggctgccac atgaggtgat aactcattat taccacagat accctatgtc 9660
caccatcctg ggtttgtcaa ttgogccgc cattgtaacc gtttccgttg cagcgtccac 9720
ctggctgttt tgcaaatcca gagtttctgt cctaactcct taccggctaa caccctaacc 9780
caggatgccg ctttgcctgg ccgtgctttg ctgcgccgc aotgcccggg ccgagaccac 9840
ctgggagtc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct 9900
gatccctctg gccgccttga ttgtagtac tcgcctgctc aagtgcgtgt gctgtgtagt 9960
gcctttttta gtcgtggccg gcgcgcgagg cgcgcgcgc taccagcacg cgaccaogat 10020
gccgagccaa gggggaatct cgtataacac catagtcaac agagcaggct acgcgccact 10080
ccctatcagc ataacaccaa caaagatcaa gctgataccc acagtgaact tggagtaoct 10140
caccctgccac tacaaaacag gaattggattc accagccatc aaatgctgcg gatctcagga 10200
atgtactcca actaacaggc ctgatgaaca gtgcaaagtc ttcacagggg ttaccogtt 10260
catgtgggga ggtgcatatt gcttttgcga cactgagaat actcaggtca gcaaggccta 10320
cgtaatgaaa tctgacgact gccttgcgga tcatgtgaa gcatacaaag cgcacacagc 10380
ctcagtgcag gcgttccctc acatcacagt gggggaacac tctattgtga ccaccgtgt 10440
tgtgaatgga gaaactcctg tgaacttcaa tggggtcaaa ctaactgcag gtccactttc 10500
cacagcttgg acaccctttg acagaaaaat cgtgcagtat gccggggaga tctataatta 10560
cgattttcct gagtatgggg caggacaacc aggagcattt ggagacatac aatccagaac 10620
agtctcaagc tcagatctgt atgccaatac caacctagt ctgcagagac ccaaagcagg 10680
agcgatccat gtgccataca ctocaggcacc atcgggtttt gagcaatgga agaaagataa 10740
agctccgtca ttgaaattca ccgccccttt cggatgcgaa atatatacaa accccattcg 10800
cgccgaaaaat tgtgctgtag ggtcaatttc attagccttt gacattcccg acgccttggt 10860
caccagggtg tcagaaacac cgacactttc agcgccgaa tgcactctta acgagtgcgt 10920
gtatccatcc gactttggcg ggatcgccac ggtaagtat tcggccagca agtcaggcaa 10980
gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac 11040
cgagcaaggg tcggcgacca ttcatcttc gaccgcaaat atccaccogg agttcaggct 11100
ccaaatatgc acatcatatg tcacgtgcaa aggtgattgt cccccccga aagaccacat 11160
tgtgacacac cccagtatc acgcccacac atttacagcc gcggtgtcaa aaaccgcgtg 11220
gacgtgggta acatccctgc tgggaggatc ggccgtaatt attataattg gcttagtgct 11280
ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa catagcagca 11340
attggcaagc tgcttatata gaacttgccg cgattggcat gccgctttaa aattttattt 11400
tattttcttt tcttttccga atcggatttt gtttttaata tttc 11444

```

<210> 2  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<220>  
 <221> modified\_base  
 <222> 11, 17  
 <223> 5-propynylcytosine

<220>  
 <221> modified\_base  
 <222> 13, 14, 16  
 <223> 5-propynyluracil

<400> 2  
 aatgctagag cguutucgca

20

<210> 3  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence



&lt;220&gt;

&lt;223&gt; Primer

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 11, 17

&lt;223&gt; 5-propynylcytosine

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 13, 14, 16

&lt;223&gt; 5-propynyluracil

&lt;400&gt; 3

aatgctagag cguutucgca

20

&lt;210&gt; 4

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 8, 14

&lt;223&gt; 5-propynylcytosine

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 10, 11, 13

&lt;223&gt; 5-propynyluracil

&lt;400&gt; 4

gctagagcgu utucgca

17

&lt;210&gt; 5

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 3

&lt;223&gt; 5-propynylcytosine

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; 10

&lt;223&gt; 5-propynyluracil

&lt;400&gt; 5

tgcgaagggu acgtcgt

17

&lt;210&gt; 6

<211> 16  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 3, 13  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 8, 9  
<223> 5-propynylcytosine

<400> 6  
tgugtgacca gaugac 16

<210> 7  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 7  
aatgctagag cgttttcgca 20

<210> 8  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 8  
aatgctagag cgttttcgca 20

<210> 9  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 9  
gctagagcgt tttcgca 17

<210> 10  
<211> 14  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 10  
tgcgaagggt acgt

14

<210> 11  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 11  
tgcgaagggt acgtcgt

17

<210> 12  
<211> 16  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 12  
tgtgtgacca gatgac

16

<210> 13  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 12, 18  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 14, 15, 17  
<223> 5-propynyluracil

<400> 13  
taatgctaga gcguutucgc a

21

<210> 14  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 12, 18  
<223> 5-propynylcytosine

<220>  
<221> modified\_base

<222> 14, 15, 17

<223> 5-propynyluracil

<400> 14

taatgctaga gcguutucgc a

21

<210> 15

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 12, 18

<223> 5-propynylcytosine

<220>

<221> modified\_base

<222> 14, 15, 17

<223> 5-propynyluracil

<400> 15

taatgctaga gcguutucgc a

21

<210> 16

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 9, 15

<223> 5-propynylcytosine

<220>

<221> modified\_base

<222> 11, 12, 14

<223> 5-propynyluracil

<400> 16

tgctagagcg uutucgca

18

<210> 17

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 4

<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 11  
<223> 5-propynyluracil

<400> 17  
ttgcgaaggg uacgt

15

<210> 18  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 4  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 11  
<223> 5-propynyluracil

<400> 18  
ttgcgaaggg uacgtcgt

18

<210> 19  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 4  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 11  
<223> 5-propynyluracil

<400> 19  
ttgcgaaggg uacgtcgt

18

<210> 20  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 4, 14

<223> 5-propynyluracil

<220>

<221> modified\_base

<222> 9, 10

<223> 5-propynylcytosine

<400> 20

ttgugtgaacc agaagac

17

<210> 21

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 21

tccatgctaa tgctagagcg ttttcgca

28

<210> 22

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 22

tgtcagttgc gaagggtacg tcgt

24

<210> 23

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 23

taatgctaga gcgttttgcg a

21

<210> 24

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 24

ttgcgaaggg tacgtcgt

18

<210> 25

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>  
<221> modified\_base  
<222> 1  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 25  
uccaatgcta gagcgttttc gca

<210> 26  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 26  
ucctgcgaag ggtacgtcgt

<210> 27  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1, 4  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 27  
uccuaatgct agagcgtttt cgca

<210> 28  
<211> 21  
<212> DNA  
<213> Artificial Sequence

23

20

24

<220>

<223> Primer

<220>

<221> modified\_base

<222> 1, 4

<223> 5-propynyluracil

<220>

<221> modified\_base

<222> 2, 3

<223> 5-propynylcytosine

<400> 28

uccutgcgaa gggtagctcg t

21

<210> 29

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 1, 4, 5

<223> 5-propynyluracil

<220>

<221> modified\_base

<222> 2, 3

<223> 5-propynylcytosine

<400> 29

uccuuaatgc tagagcggtt togca

25

<210> 30

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 1, 4, 5

<223> 5-propynyluracil

<220>

<221> modified\_base

<222> 2, 3

<223> 5-propynylcytosine

<400> 30

uccuutgcga agggtagctc gt

22

<210> 31

<211> 26



<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 31  
tccttcaatg ctagagcggtt ttcgca 26

<210> 32  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 32  
tccttctgcg aagggtacgt cgt 23

<210> 33  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 33  
tgccagctac actgtgacgac cagatgac 28

<210> 34  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 34  
tattgtcagt tgccaagggt acgt 24

<210> 35  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 35  
tattgtcagt tgcgacgggt acgt 24

<210> 36  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 36  
tctatagtca gttgacgacgg gtacgt 26

<210> 37  
<211> 31  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 37  
tgtcagctac attgtgtgac caaatgactg g 31

<210> 38  
<211> 31  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 38  
taccagccac actttgcat cagatgacag g 31

<210> 39  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 39  
tccatgctaa cgccagagcg ttttcgca 28

<210> 40  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 40  
tccatgctaa cgccagagcg ttttcgca 28

<210> 41

<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 41  
tgacgtagac cccagagtc cgttt 25

<210> 42

<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 42  
tggcgctatg atgaaatctg gaatggt

27

<210> 43  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 43  
tggcgctatg atgaaatctg gaatggt

27

<210> 44  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 44  
tgccttcacg ggcgatgaca acat

24

<210> 45  
<211> 31  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 45  
tgtcgccga ggattttgat gctatcatag c

31

<210> 46  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 46  
tgcggtaccg tcaccatttc agaacac

27

<210> 47  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 7, 8, 14  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 11, 13  
<223> 5-propynyluracil

<400> 47  
gcacttccaa uguccaggat

20

<210> 48  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 7, 8, 14  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 11, 13  
<223> 5-propynyluracil

<400> 48  
gcacttccaa uguctaggat

20

<210> 49  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 5, 7, 10, 11  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 8, 14, 16  
<223> 5-propynyluracil

<400> 49  
ggcgcacutc caauguc

17

<210> 50  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 4, 16, 17

<223> 5-propynylcytosine

<400> 50

ttgcagcaca agaatccctc

20

<210> 51

<211> 14

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 4, 5

<223> 5-propynyluracil

<220>

<221> modified\_base

<222> 9, 10, 11

<223> 5-propynylcytosine

<400> 51

tggugagcc caac

14

<210> 52

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 52

gcacttccaa tgtccaggat

20

<210> 53

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 53

gcacttccaa tgtctaggat

20

<210> 54

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 54  
ggcgcacttc caatgtc 17

<210> 55  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 55  
ttgcagcaca agaatccctc 20

<210> 56  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 56  
ttgcagcaca agaatccctc 20

<210> 57  
<211> 14  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 57  
tggttgagcc caac 14

<210> 58  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 8, 9, 15  
<223> 5-propynylcytosine

<220>  
<221> modified\_base  
<222> 12, 14  
<223> 5-propynyluracil

<400> 58  
tgcacttcca auguccagga t 21

<210> 59  
<211> 21  
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 8, 9, 15

<223> 5-propynylcytosine

<220>

<221> modified\_base

<222> 12, 14

<223> 5-propynyluracil

<400> 59

tgcaattcca auguccagga t

21

<210> 60

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 8, 9, 15

<223> 5-propynylcytosine

<220>

<221> modified\_base

<222> 12, 14

<223> 5-propynyluracil

<400> 60

tgcaattcca auguctagga t

21

<210> 61

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> modified\_base

<222> 6, 8, 11, 12

<223> 5-propynylcytosine

<220>

<221> modified\_base

<222> 9, 15, 17

<223> 5-propynyluracil

<400> 61

tggcgacut ccaauguc

18

<210> 62  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 5, 17, 18  
<223> 5-propynylcytosine

<400> 62  
tttgcagcac aagaatccct c 21

<210> 63  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 5, 17, 18  
<223> 5-propynylcytosine

<400> 63  
tttgcagcac aagaatccct c 21

<210> 64  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 5, 17, 18  
<223> 5-propynylcytosine

<400> 64  
tttgcagcac aagaatccct c 21

<210> 65  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 5, 6  
<223> 5-propynyluracil



<220>  
<221> modified\_base  
<222> 10, 11, 12  
<223> 5-propynylcytosine

<400> 65  
ttgguugagc ccaac 15

<210> 66  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 66  
tgggcactt ccaatgtcca ggat 24

<210> 67  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 67  
tctgtcactt tgcagcacia gaatccctc 29

<210> 68  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 68  
tgcacttcca atgtccagga t 21

<210> 69  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 69  
tttgcagcac aagaatccct c 21

<210> 70  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>

<221> modified\_base  
<222> 1  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 70  
uccgcacttc caatgtccag gat

23

<210> 71  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 71  
uccttgacgc acaagaatcc ctc

23

<210> 72  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1, 4  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 72  
uccugcactt ccaatgtcca ggat

24

<210> 73  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1, 4  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 73  
uccuttgcag cacaagaatc cctc

24

<210> 74  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1, 4, 5  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 74  
uccuugcact tccaatgtcc aggat

25

<210> 75  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<220>  
<221> modified\_base  
<222> 1, 4, 5  
<223> 5-propynyluracil

<220>  
<221> modified\_base  
<222> 2, 3  
<223> 5-propynylcytosine

<400> 75  
uccuuttgca gcacaagaat ccctc

25

<210> 76  
<211> 26  
<212> DNA  
<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 76

tccttcgcac ttccaatgtc caggat

26

&lt;210&gt; 77

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 77

tccttcttgc agcacaagaa tccctc

26

&lt;210&gt; 78

&lt;211&gt; 28

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 78

tgacgactat ccgctggttg agcccaac

28

&lt;210&gt; 79

&lt;211&gt; 27

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 79

tgtcactttg caacacaaga atccctc

27

&lt;210&gt; 80

&lt;211&gt; 27

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 80

tgtcactttg caacacaaga atccctc

27

&lt;210&gt; 81

&lt;211&gt; 27

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 81

tgtcactttg cagcacaaga atccctc

27

<210> 82  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 82  
tgacgactat ccgctggttg agcccaac

28

<210> 83  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 83  
tgacgactat ccgctggttg agcccaac

28

<210> 84  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 84  
tgctggtgca cttccaatat ccaggat

27

<210> 85  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 85  
tgccggtgcg ctgcctatgt ccaa

24

<210> 86  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 86  
tcgctctggc attagcatgg tcatt

25

<210> 87  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<400> 87

tatgttggtcg tcgccgatga acgc

24

<210> 88

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 88

tacgatgttg tcgtcgccga tgaa

24

<210> 89

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 89

tccaagtggc gcacctgtct gccat

25

<210> 90

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 90

tcaccttggc ttttgtcaaa ggaggc

26

<210> 91

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 91

tggtagttct ctcatattgtg tgacgttgca

30

<210> 92

<211> 4108

<212> DNA

<213> Artificial Sequence

<220>

<223> Calibrant sequence pVIR001

<400> 92

agcgccaat acgcaaaccg cctctccccg cgcgttgcc gattcattaa tgcagctggc 60  
acgacagggt tcccgaactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120  
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180

ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac	gccaaactat	240
ttaggtagcg	cgttagaata	ctcaagctat	gcatacaagct	tggtagcgag	ctcgatcca	300
ctagtaacgg	ccgccagtgt	gctggaattc	aggactcaag	ctggagtgtg	gaatgcatgc	360
ttattcacat	acaaactacc	accatcacag	cctggtaagt	tcaagtttga	caagactctt	420
gaacctacac	acaattgcat	tggctgggta	acgatcaacg	ttacaattcc	aaaacaaaca	480
aacaccatca	gtgaatttat	cctctgtcac	atcttggtata	atcccaacc	ataaggtgtg	540
gagttttctac	atcactgtaa	atttaacata	ttatgccagc	cacogtaaaa	cttgcttgtt	600
ccatgacgac	tatgogctgg	ttgagcccaa	ccagcagttt	ttgogcgctg	tccgcactga	660
cttgccagta	tgccagtcac	ttgggtcacac	aatgtagctg	ggtctgtcac	tttgacgac	720
aagaatccct	cgcggtgcac	cgtagcagca	tagcctgaag	gcttcccata	caggcctgga	780
agctattctt	ttaacgacgt	acccgtcgca	actaacata	ctgcggcgcg	gcgcacttcc	840
aatgtcaagg	atcgtgtcgg	atgggtccac	ctccgttcag	ttttgaagcc	agatgcgaaa	900
acgctctggc	attagcatgg	tcattttgtc	ctgaattctg	cagatatcca	tcacactggc	960
ggcgcctcga	gcattgcatt	agaggcccca	attcgcccta	tagtgagctg	tattacaatt	1020
cactggccgt	cgttttacaa	cgtcgtgact	gggaaaacc	tggcgttacc	caacttaate	1080

gccttgacgc	acatccccct	ttcgccagct	ggcgtaatag	cgaagaggcc	cgcaccgatc	1140
gcccttccca	acagttgcgc	agcctatacg	tacggcagtt	taaggtttac	acctataaaa	1200
gagagagcgt	ttatcgtctg	tttgtggatg	tacagagtga	tattattgac	acgcggggcg	1260
gacggatggg	gatccccctg	gccagtgcac	gtctgctgtc	agataaagtc	tcccgtagaac	1320
tttaccgggt	gggtgcatac	ggggatgaaa	gctggcgcat	gatgaccacc	gatatggcca	1380
gtgtgccggg	ctccgttata	ggggaagaag	tggctgatct	cagccaccgc	gaaaatgaca	1440
tcaaaaacgc	cattaacctg	atgttctggg	gaatataaat	gtcaggcatg	agattatcaa	1500
aaaggatctt	cacctagatc	cttttccact	agaaagccag	tccgcagaaa	cgggtgctgac	1560
cccgatgaa	tgtagcctac	tgggctatct	ggacaaggga	aaacgcaagc	gcaaagagaa	1620
agcaggtagc	ttgcagtggg	cttacatggc	gatagctaga	ctgggcggtt	ttatggacag	1680
caagcgaacc	ggaattgcc	gctggggcgc	cctctggtaa	ggttgggaag	ccctgcaaag	1740
taaactggat	ggctttcttg	ccgccaagga	tctgatggcg	caggggatca	agctctgac	1800
aagagacag	atgagcatcg	tttcgcata	ttgaacaaga	tggattgcac	gcaggttctc	1860
cggcgcgttg	gggtggagag	ctattccggt	atgactgggc	acaacagaca	atcggtctgt	1920
ctgatgccgc	cgtgttcogg	ctgtcagcgc	aggggcgcgc	ggttcttttt	gtcaagaccg	1980
acctgtccgg	tgcctgaat	gaactgcaag	acgaggcagc	gcggtctatc	tggctggcca	2040
cgcggggcgt	tccttgccga	gctgtgctcg	acgttgtcac	tgaagcggga	agggactggc	2100
tgctattggg	cgaagtgcgc	gggcaggatc	tcctgtcatc	tcaccttgct	cctgccgaga	2160
aagtatccat	catggctgat	gcaatgcggc	ggctgcatac	gcttgatccg	gctacctgcc	2220
cattcgacca	ccaagcgaaa	catcgcatcg	agcgagcacg	tactcggtat	gaagccggtc	2280
ttgtcgacca	ggatgatctg	gacgaagagc	atcaggggct	cgcgcagccg	gaactgttcg	2340
ccaggctcaa	ggcgagcatg	ccgcagcgcg	aggatctcgt	cgtgacctat	ggcgatgctt	2400
gcttgccgaa	tatcatgggt	gaaaatggcc	gcttttcttg	attcatcgac	tgtggccggc	2460
tgggtgtggc	ggaccgctat	caggacatag	cgttggtctac	ccgtgatatt	gctgaagagc	2520
ttggcggcga	atgggctgac	cgttccctcg	tgttttacgg	tatcgccgct	cccgattcgc	2580
agcgcatcgc	cttctatcgc	cttcttgacg	agttcttctg	aattattaac	gcttacaatt	2640
tcctgatcgc	gtattttctc	cttacgcata	tgtgcggtat	ttcacacccg	atcaggtggc	2700
acttttccgg	gaaatgtgcg	cggaaacctt	atttgtttat	ttttctaaat	acattcaaat	2760
atgtatccgc	tcattgagaca	ataaccttga	taaatgcttc	aataatagca	cgtgaggagg	2820
gccaccatgg	ccaagttgac	cagtgcggtt	ccggtgctca	cgcgcgcgca	cgtcgccgga	2880
gcggctcag	tctggaccga	ccggctcggg	ttctcccggg	acttcgtgga	ggacgacttc	2940
gccgggtgtg	tcggggacga	cgtgacctg	ttcatcagcg	cgggtccagga	ccagggtggtg	3000

ccggacaaca	ccctggcctg	ggtgtgggtg	cgcggcctgg	acgagctgta	cgcaggtggg	3060
tcggaggtcg	tgtccacgaa	cttccgggac	gcctccgggc	cggccatgac	cagatcggc	3120
gagcagccgt	gggggcggga	gttcgcctcg	cgcgaccggg	ccggcaactg	cgtgcacttc	3180
gtggccgagg	agcaggactg	acacgtgcta	aaacttcatt	tttaatttaa	aaggatctag	3240
gtgaagatcc	tttttgataa	tctcatgacc	aaaatccctt	aacgtgagtt	ttcgttccac	3300
tgagcgtcag	accccgtaga	aaagatcaaa	ggactctctt	gagatccctt	ttttctgcgc	3360
gtaatctgct	gcttgcaaac	aaaaaaacca	cgcctaccag	cgggtggtttg	tttgccggat	3420
caagagctac	caactctttt	tccgaaggta	actggcttca	gcagagcgca	gataccaaat	3480
actgttcttc	tagttagacc	gtagttaggc	caccacttca	agaactctgt	agcaccgcct	3540
acatacctcg	ctctgctaatt	cctgttaacca	gtggctgctg	ccagtggcga	taagtcgtgt	3600
cttaccgggt	tggactcaag	acgatagtta	ccggataagg	cgcagcggtc	gggctgaacg	3660
gggggttcgt	gcacacagcc	cagcttgag	cgaacgaact	acaccgaact	gagataccta	3720

```
cagcgtgagc tatgagaaaag cgccacgctt cccgaaggga gaaaggcgga caggtatccg 3780
gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg aaacgcctgg 3840
tatctttata gtcctgtcgg gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc 3900
tcgtcagggg ggcgagagcct atggaaaaac gccagcaacg cggccttttt acggttcctg 3960
gccttttgcg ggccttttgc tcacatgttc tttcctgcgt tatccctga ttctgtggat 4020
aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac gaccgagcgc 4080
agcgagtcag tgagcgagga agcggaag 4108
```